Information theoretic approaches to functional neuroimaging

Dirk Ostwald\textsuperscript{a,b,c,⁎}, Andrew P. Bagshaw\textsuperscript{a,b}

\textsuperscript{a}School of Psychology, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK
\textsuperscript{b}Birmingham University Imaging Centre, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK
\textsuperscript{c}Department of Neurology and Bernstein Center for Computational Neuroscience, Charité, 10117 Berlin, Germany

Received 3 March 2011; revised 17 June 2011; accepted 6 July 2011

Abstract

Information theory is a probabilistic framework that allows the quantification of statistical non-independence between signals of interest. In contrast to other methods used for this purpose, it is model free, i.e., it makes no assumption about the functional form of the statistical dependence or the underlying probability distributions. It thus has the potential to unveil important signal characteristics overlooked by classical data analysis techniques. In this review, we discuss how information theoretic concepts have been applied to the analysis of functional brain imaging data such as functional magnetic resonance imaging and magneto/electroencephalography. We review studies from a number of imaging domains, including the investigation of the brain’s functional specialization and integration, neurovascular coupling and multimodal imaging. We demonstrate how information theoretical concepts can be used to answer neurobiological questions and discuss their limitations as well as possible future developments of the framework to advance our understanding of brain function.

Keywords: Information theory; Functional neuroimaging

1. Introduction

Information theory is a branch of probability theory and has its origin in the study of communication [1]. Mathematically, information theory derives from principles developed in thermodynamics, and because of their generality, information theoretical concepts can be applied to various forms of signal processing [2]. In this review, we discuss how information theoretic concepts can be applied to non-invasive functional brain imaging data such as functional magnetic resonance imaging (fMRI) and magneto/electroencephalography (M/EEG). Applications of information theoretical concepts to functional brain imaging data are increasingly common and represent one of the more principled approaches to functional brain imaging data analyses owing to the relative simplicity of the assumptions underlying the framework. Information theoretic analyses also benefit from an explicit and transferrable mathematical definition of “information,” which is advantageous given that the most common metaphor for the brain is that of a dynamic information processing device [3]. Furthermore, it can be shown that there are deep connections between commonly employed data analysis methods in the functional brain imaging field, such as those derived from classical statistics (e.g., variants of the general linear model) and machine learning (e.g., multivariate classification techniques), and information theoretical concepts. These formal relationships demonstrate that most data analysis methods can be viewed as information theoretic approaches that embody certain regularization constraints such as specific forms of statistical dependencies and signal distributions [4].

Information theoretic approaches to neural data analysis have been pioneered in the application to invasive electrophysiological recordings [5] and can be considered an established signal processing methodology in electrophysiology [6]. The application of information theoretic approaches to functional imaging data has so far received far less attention. However, as both fields face similar questions (e.g., Which aspect of the recorded data is informative about the stimulus? How do correlations between signals shape the informativeness of the overall response?), the exploration of information theoretic approaches to functional imaging is a worthwhile endeavour.

⁎ Corresponding author.
E-mail address: dirk.ostwald@bccn-berlin.de (D. Ostwald).
for increasing the cross-talk between the disciplines and placing them on a similar analytical footing.

After briefly introducing the main concepts of information theory (Sections 2 and 3), we aim to elucidate how information theoretic concepts have been applied in a variety of brain imaging contexts: specifically, we review applications of information theory to the study of functional specialization (Sections 4 and 5), functional integration (Section 6), neurovascular coupling (Section 7) and multimodal imaging (Section 8).

2. Information theoretic concepts and empirical information decomposition

Three quantities are of central importance in information theory: entropy, conditional entropy and mutual information. The entropy

\[ H(X) = - \sum_x p(x) \log_2 p(x) \]  

(1)

of a signal \( X \) is a measure of its uncertainty and becomes maximal for a uniformly distributed signal. Mathematically, the entropy of a signal is a functional (i.e., a function of a function) of its probability density or mass function. The conditional entropy \( H(X|Y) \) of a signal \( X \) is the entropy that remains after the variability due to another signal \( Y \) has been accounted for:

\[ H(X|Y) = - \sum_{x,y} p(x,y) \log_2 p(x|y) \]  

(2)

The mutual information \( I(X;Y) \) between two signals \( X \) and \( Y \) can then be expressed as the difference of the entropy in signal \( X \) and its conditional entropy given \( Y \), i.e., it represents the remaining entropy of signal \( X \) after all variability of \( X \) associated with \( Y \) has been accounted for:

\[ I(X;Y) = H(X) - H(X|Y) \]  

(3)

Equivalently, the mutual information between two variables can be viewed as a measure of their statistical non-independence. It can be shown that the entropy difference discussed above is equivalent to the Kullback-Leibler divergence between the joint distribution of two variables and its factorized equivalent [2]. That is, mutual information is a measure of the difference between the actual joint probability distribution and that which would be observed if the variables were independent. It should be noted that statistical independence is a singular phenomenon, in the sense that if two signals \( X \) and \( Y \) are not independent, they can be dependent in an arbitrary number of ways. The use of entropy as a measure of a signal’s variability, as well as the use of mutual information as a measure of the non-independence of two or more signals, is hence motivated by their model-free character: unlike other measures of signal covariation (e.g., the correlation coefficient or the general linear model), the mutual information of two signals does not imply a linear relationship between them. It hence has the potential to unveil important signal relationships in the absence of linearity and parametric assumptions about the residual distribution (e.g., Gaussian noise) while remaining as sensitive to affine relationships as standard methods based on these assumptions (Fig. 1).

Mutual information quantities can be computed between any kind of experimental signals, i.e. both \( X \) and \( Y \) can both represent observed signals such as the fMRI signal from different voxels, or one of them a probabilistic stimulus variable and the other an empirical signal. Moreover, as the conditional entropy can also be computed for multiple conditioning variables, multivariate extensions of mutual information such as the information that two variables convey about a third, \( I(X;Y,Z) \), are readily conceivable.

Evaluating the mutual information from a set of variables about another variable is the central theme of information decomposition schemes developed in the analysis of electrophysiological data [7,8]. Four concepts are central in this context: noise and signal correlations, information redundancy and information synergy [9]. Informally, noise correlations refer to signal co-variation at a fixed value of the experimental variable, e.g., in the absence of differential stimulation, while signal correlations refer to co-variations between response signals solely induced by differential stimulation, in the absence of noise correlation. Information synergy refers to the additional information the observation of the joint distribution of a set of signal variables conveys about an experimental variable, beyond the information implicit in the respective marginal signal distributions. Negative values of information synergy are referred to as information redundancy and represent the situation when multiple response variables convey the same information. Information decomposition concepts provide useful extensions to common approaches to functional data analysis as will be discussed below, for example, in relation to the analysis of multivariate and multimodal imaging data.

3. Estimating entropy and mutual information from empirical data

The first step in the application of information theoretic concepts to empirical data is feature selection, i.e., the selection of the data characteristic whose probability distributions one would like to examine by quantification of that feature for each experimental trial. While exhaustive feature searches are at least conceptually possible, feature selection is usually guided by the experimenter’s intuition and relevant background literature. It should be noted that the information theoretic framework requires the imaging data acquired on a single trial to be collapsed onto a single (potentially multivariate) data point [e.g., the maximum of the haemodynamic response function (HRF)]. This dimensionality reduction is commonly based on assumptions about the
potential relevance of different aspects of the data (e.g., time- vs. frequency-domain features) and, hence, always entails the experimenter’s theory about the neurobiological origin and significance of the signal. While previous applications of the framework commonly relied on in the extraction of univariate single-trial data points directly from the sampled raw data, feature extraction could potentially be augmented using analytical models of the single-trial response. Fitting these single-trial models (e.g., of the HRF for fMRI or the evoked potential for M/EEG) to observed data and using single-trial parameter estimates of the ensuing model-fit for subsequent information theoretic analyses might potentially allow a more principled and neurobiologically-grounded data analysis approach with higher signal-to-noise ratio.

After the relevant features have been decided upon, their single-trial values form the basis for the computation of information theoretic quantities. As discussed above, information theoretic quantities are functionals of probability mass or density functions. Estimation of mutual information hence requires, at least in principle, the estimation of the entire joint probability distribution of a given set of experimental variables. This is a marked difference to methods in classical statistics, where estimators such as the sample mean and sample standard deviation are typically functions of the realizations of random variables, but not of the corresponding probability density functions. The fundamental method to determine an empirical signal’s probability distribution is a histogram analysis. However, as histogram analyses can only approximate the true underlying probability distribution, the approximation error or estimation bias has to be taken into account.

The estimation bias problem for entropy and mutual information has received considerable attention in the statistical and electrophysiological literature, and a large number of different methods has been developed for its correction (see Ref. [10] for a review). The sampling bias of a statistic is dependent on a number of factors, including the number of trials for which it is evaluated, the properties of the true underlying probability distribution as well as the properties of the estimator itself, for example, the free parameters in a histogram analysis (number of bins and bin size). Bias correction procedures commonly employ some parametric or nonparametric constraints on the (assumed) true probability distributions and, based on analytical results, provide quantities that can be subtracted from the estimated mutual information. Prevalent bias correction techniques in...
neurophysiological data analyses are Panzeri-Treves (PT) correction, quadratic extrapolation (QE), best universal bound estimation (BUB), Nemenman-Schafee-Bialek and shuffling correction [10], which have recently become available as a Matlab toolbox [9]. A non-parametric bias correction method increasingly used in imaging data analysis is that developed by Kraskov et al. [11]. The optimal bias correction methods for information theoretic analyses of brain imaging data have not yet been thoroughly examined. The results of information theoretic analyses hence require careful interpretation and, if possible, validation using different approaches to information estimation as well as reproduction in independent data sets.

An example of an information theoretic analysis of functional imaging data is provided as Fig. 2. The data presented is from a single-subject performing a simultaneous EEG-fMRI experiment (for experimental details, please see Ref. [38]). Briefly, the subject was visually presented with checkerboards of high and low contrast (85 trials per contrast condition), while simultaneously, EEG and fMRI responses were monitored. After standard EEG-fMRI data pre-processing, one data feature from each modality was quantified on each experimental trial. Based on previous research on the contrast sensitivity of visual cortex, for the EEG, the P100 amplitude of a contralateral occipital electrode and for the fMRI the maximum of the post-stimulus haemodynamic response from contra-lateral primary visual cortex were extracted. The combined data for both high-contrast (S1) and low contrast (S2) trials is plotted in Panel 1 of Fig. 2.

After single-trial feature selection, the next step towards an information theoretic analysis is the evaluation of the joint stimulus-response probability distribution. As discussed above, this can be carried out non-parametrically or parametrically. A non-parametric histogram analysis with optimized binning parameters (see Ref. [38] for details) of the data is shown in Panel 2 (right) of Fig. 2. On the left, the results of a parametric approach, namely, the Gaussian method [9], are presented. In brief, the Gaussian method entails the estimation of the stimulus-independent and stimulusconditional data covariances.

After the joint stimulus response probability mass or density functions have been estimated, the analysis proceeds by evaluating the information theoretic quantities of interest (Fig. 2, Panel 3). Usually, upon evaluation of the information of the joint response about the stimulus, some form of information breakdown scheme is employed to gain further insight into the role of signal and noise correlations. As an example, we here consider the information synergy, a measure of the information the joint response distribution can provide in addition to the sum of the marginal response distributions. Also in this step, the information theoretic quantities should be corrected for the limited sampling bias. Here, we employed PT and shuffling correction for the nonparametric information estimation and the analytically determined bias subtraction embodied in the Gaussian method for the parametric information estimation.

The last step of an information theoretic analysis is the interpretation of the results. Based on the non-parametric and parametric approaches, for the current data, the following conclusion can be drawn: both methods agree on the fact that the information estimates for the marginal distributions of each EEG and fMRI are non-zero and larger for the EEG than for the fMRI. Further, the information about the stimulus contrast contained in the joint distribution of both EEG and fMRI features is larger than the information provided by the marginal distributions. Finally, the methods diverge for the synergy term: while the non-parametric method advocates positive information synergy, the parametric method advocates negative synergy (redundancy). This difference is best explained by the shortcomings of each method with respect to the data: while the histogram binning necessarily removes some of the information present, the Gaussian method assumes that the data are indeed sampled from Gaussian distributions. Slight mismatches in the data—methodology relationship can hence lead to different interpretations of the results, highlighting the fact that optimisation of the methodology for the particular type of data under consideration is crucial. In conclusion, for this data, there is good evidence for more pronounced information in the EEG feature than in the fMRI feature and the joint with respect to the marginal distribution, but no conclusion can be drawn on whether the data features are in fact synergistic or redundant.

4. Applications in univariate studies of functional specialization

The traditional aim of functional brain imaging is to establish structure-function relationships by studying the functional specialization of brain areas [12]. The prevalent method used for this purpose is a mass-univariate (i.e., voxel-wise) application of the general linear model (GLM). GLM-based fMRI data analysis is an established methodology that can successfully be used in a variety of contexts. Nevertheless, the GLM approach embodies a number of model-assumptions (e.g., shape of the impulse HRF, linearity, Gaussian residuals). Information theoretical methods applied to univariate fMRI signal features offer a model-free alternative to the study of functional specialization, which has been explored in a number or recent studies.

In one of the first studies, De Ajauro et al. [13] proposed a method for the analysis of event-related fMRI time series based on entropy estimation as a measure of the blood oxygenation level-dependent (BOLD) signal’s evoked single-trial variability. To demonstrate their approach, the method was applied to fMRI data acquired under visual and motor simulation. Specifically, each voxel’s post-stimulus time-window was divided into two segments, the first of which contained the event-related signal increase and the second the baseline signal. For each time-window, a histogram analysis was used, which discretized the possible
1 - Single Trial Feature Selection

\[ R_1 \text{ - EEG P100 Amplitude (\(\mu V\))} \]

\[ R_2 \text{ - fMRI HRF Maximum (\%) } \]

2 - Estimation of probability functions

Nonparametric Estimation

Parametric Estimation

3 - Evaluation of information theoretic quantities and bias correction

\[ I(S; R_1, R_2) = \sum_{x \in S} \sum_{r_i \in R_i} \sum_{s \in S} p(s, r_i, r_2) \log_2 \left( \frac{p(s, r_i, r_2)}{p(s) p(r_i)} \right) \]

\[ I(S; R_i) = \sum_{x \in S} \sum_{r_i \in R_i} p(s, r_i) \log_2 \left( \frac{p(s, r_i)}{p(s) p(r_i)} \right) \quad i = 1, 2 \]

+ Bias Estimate Subtraction

\[ H(R) = \frac{1}{2} \log_2 (2\pi e |\Sigma(R)|) \]

\[ H(R|S) = \frac{1}{2} \sum_s p(s) \log_2 \left( (2\pi e |\Sigma_s(R)|) \right) \]

+ Analytical Bias Subtraction

Information Decomposition: Synergy = \( I(S; R_1, R_2) - (I(S; R_1) + I(S; R_2)) \)

4 - Interpretation

Fig. 2. An information theoretic data analysis example. This figure demonstrates the process of applying information theoretic concepts to experimental functional imaging data using single-subject data from a simultaneous EEG-fMRI experiment [10]. For a detailed discussion, please refer to the main text. On Panels 2–4, the left hand items correspond to a nonparametric approach to information estimation, while the right hand items correspond to a parametric approach.
signal values and pooled (not averaged) the time-wise measurements to determine the probability of the signal assuming a specific value. Based on these estimated probability mass functions, the entropy of the signal in each time-window was computed. To create a statistical map based on the entropy time-courses for each voxel, the cross-correlation between the entropy signal and a simulated sawtooth function was used. The results of this approach demonstrated good topographical correspondence between areas labelled activated by the entropy method and a classical cross-correlation method using a lagged gamma function as a model for the HRF. The authors noted that, while being model free, the approach still required an implicit model of the HRF which was used in the determination of the two time-windows.

Using a series of different experimental paradigms (motor-learning, audio, visual and audio-visual), Fuhrmann-Alpert et al. [14,15] provided a comprehensive demonstration of the use of information theoretic concepts in fMRI data analysis. Three questions were central to the work of Fuhrmann-Alpert et al.: (a) Which voxels are most informative about a preceding stimulus condition? (b) At what post-stimulus time are the voxels most informative about the stimulus condition? and (c) How informative is a set of voxels (region-of-interest) with respect to another set of voxels irrespective of the timings of the experimental paradigm? To gain insight into these questions, the authors developed the following data analysis scheme: to study spatial specificity of the information content, the stimulus conditional entropy of the BOLD response was calculated for each voxel and each post-stimulus latency, i.e., for each sampling point. Likewise, the stimulus unconditional entropy of the BOLD response was computed by pooling all sampling data across the entire time-series. Entropy estimation was then achieved by discretization of the voxel’s response values with a constant bin size for both conditional and unconditional entropies, histogram-based estimation of the probability mass function and subsequent evaluation of the respective entropies. To create spatial mutual information maps, the latency of maximal informativeness was determined for each voxel, i.e., the mutual information values were displayed for different post-stimulus latencies. In order to threshold these maps, a shuffling approach was employed to determine a mutual information significance threshold based on an empirical null distribution. To address the question of temporal maximal informativeness, the authors plotted the post-stimulus latency for each voxel that maximized its information content in map format. Finally, in order to use mutual information as a measure of the functional connectivity between a pair of pre-selected regions of interest, the unconditional entropy of a first region of interest and its conditional entropy, conditioned on each of six possible response values of a second region of interest, were computed.

Using these three approaches, in Ref. [14], the authors provided evidence that during motor learning, task-related information in unimodal motor cortical areas precedes the response in higher-order multimodal association areas, including posterior parietal cortex. Moreover, prefrontal brain areas associated with reduced activity during motor learning were found to be informative about the task at later time points. In Ref. [15], it was demonstrated that after audio-visual stimulation, the earliest informative activity occurs in right Heschl’s gyrus, left primary visual cortex and the posterior portion of the superior temporal gyrus. At a subsequent latency, informative activity was detected in the anterior portion of the superior temporal gyrus, middle temporal gyrus, right occipital cortex and inferior frontal cortex (Fig. 3A). Interestingly, simultaneous audio-visual stimulus presentation resulted in shorter latencies in multiple cortical areas compared with unimodal auditory or visual stimulation.

In summary, the application of information theoretic measures to fMRI signal-processing in the study of univariate functional specialization has been demonstrated to yield neurobiologically meaningful results. Nevertheless, what has not been directly demonstrated is how an information theoretic approach might outperform traditional mass-univariate GLM analyses. For example, what is missing so far is an explicit demonstration of the information theoretic framework’s sensitivity to fMRI signal non-linearities, which would be missed using traditional GLM analyses and might be of neurobiological importance.

5. Information theoretic approaches to multivariate studies of functional specialization

In addition to the traditional mass-univariate GLM approach to the study of functional specialization, multi-voxel pattern analyses (MVPA) of fMRI data have become popular during the last decade [16–18]. MVPA analyses are motivated by the intuition that individual voxels might only display small functional biases, whereas the joint activity of a voxel collective might convey strong functional information [19,20]. In principle, the classification accuracy that a given statistical model can achieve in a cross-validation-based MVPA and the mutual information the respective voxel activation profile conveys about the experimental variable are directly equivalent [4]. Moreover, estimating the mutual information for a given voxel activation pattern instead of classification accuracy offers a more principled approach to identify informative data aspects and potentially enables better between-studies comparison [6]. The most comprehensive study that investigated information theoretic approaches to MVPA is that by Rolls et al. [21], reviewed below. Further studies that integrated information theoretical concepts and MVPA are for example the studies by Kriegerkorte et al. [22] and Pessoa et al. [23].

In the study by Rolls et al. [21] the information theoretic approach was used to measure how well activations of a given set of voxels in a pre-determined region of interest can
predict or discriminate between the affective state of the subject, i.e., how well the voxel pattern activations differ between different subjective judgements of a thermal stimulus. To estimate the information values a decoding approach developed previously by the same authors [24] was used to optimally cope with the undersampled response space. The central idea of this approach is to compute the joint distribution of stimulus and predicted stimulus $S', p(S,S')$. This distribution is spanned by the product of the dimensionality of the stimulus space, and not by the product of the dimensionality of the stimulus and response space and is hence less undersampled for high dimensional response spaces. The predicted stimulus was obtained by comparing the multi-voxel activation pattern on a given test to a training pattern obtained by evaluating the activations to known stimuli. To account for the limited sampling bias, the estimated
information values were then corrected using PT bias correction [25]. Using this approach, it was found that the subjective pleasantness of the stimulus could be predicted with 60–80% accuracy from the activations of voxels located in orbitofrontal medial prefrontal and pregenual cingulate cortex, with an information equivalent of 0.1–0.2 bits (with a theoretically possible maximum of 0.6 bits in this case). Further, the information estimates/decoding accuracies were typically higher with multiple voxels compared to one voxel. The information estimates increased sublinearly with the number of voxels, indicating redundant information representation between voxels. These redundancy effects were found even if the analyzed voxel data originated from regions of interest in different brain areas. By comparison of their results with single-cell physiology data, the authors concluded that the integrating property of the fMRI signal potentially reduces information represented by small populations of neurons.

As demonstrated by this study, in addition to providing insight into the question of whether a given voxel activation pattern is informative about an experimental variable, the focus on signal and noise correlations that is implicit in the information theoretic framework allows the influence of voxel-by-voxel covariation in information encoding to be assessed directly. This spatial covariance, especially when sensitive to different experimental contexts, might provide important evidence about the neural code that is discarded if the data-analytical focus is restricted to overall classification accuracy. However, while the information theoretic approach provides a richer and more principled approach to MVPA than the frameworks currently employed, it should be noted that advanced machine learning algorithms provide excellent generalizability from training to test data and, hence, might be more sensitive to pattern information than current approaches to the estimation of mutual information. Future methodological studies might address questions of both sensitivity and specificity for the estimation of information based on direct approaches as discussed here and classification-algorithm based approaches for MVPA as pioneered in Ref. [21]. A formal treatment of the relationship between cross-validation based classification accuracy for a given classification algorithm and mutual information would be a first step to identify potential trade-offs in applying both approaches to fMRI-MVPA.

6. Applications to functional integration

So far the use of information theoretic concepts to study the brain’s functional specialization has been discussed. However, the study of functional integration is of equal importance to our understanding of the neurobiological underpinnings of brain function and requires techniques to address the issue of how multiple distributed brain regions communicate in order to produce a functioning network [12]. Placed in these terms, the suitability of information theory, a tool specifically developed for the study of network communication, seems clear. Although mutual information itself does not deal with the directionality or time-varying nature of the link between systems, advanced information theoretic measures are available to address these questions. For example, transfer entropy [26] is a model-free measure of directed (time-asymmetric) information transfer which bears equivalence to more commonly used metrics such as Granger causality under certain assumptions. Two recent studies used the estimation of transfer entropy in the analysis of univariate [27] and multivariate fMRI data [28].

Hinrichs et al. [27] focussed on information transfer between activated visual cortical areas in a right homonymous hemianopia patient. Upon establishing that transfer entropy successfully recovers linear and nonlinear couplings for simulated paired time-series, the authors used a kernel estimation procedure to determine information transfer measures for different types of visual stimulation. Specifically, stimulation by visual movement elicited high information flow from striate cortex to area V5 and stimulation by color change resulted in high information flow from striate cortex to area V4/V8 in the patient’s healthy hemisphere. Moreover, for the lesioned hemisphere, both kinds of visual stimulation resulted in stronger flow between high level visual areas V5 and V4/V8 as well as top-down flow from V4/V8 to V2 providing evidence for the role of subcortical pathways bypassing primary visual cortex in blind-sight patients.

While the study by Hinrichs et al. [27] focussed on time-courses reflecting the average activity of regions-of-interest, Lizier et al. [28] demonstrated the use of information theory in the study of multivariate information transfer. Using multivariate transfer entropy, the authors thus provided a novel MVPA approach to interregional connectivity analysis. Using the same non-parametric approach to entropy estimation as Hinrichs et al. [27], the authors applied their method to fMRI data from a visuo-motor tracking task and were able to recover a tiered structure of top-down control for manual tracking. In this structure, the top two tiers correspond to the pre-motor cortical network including primary motor cortex, the supplementary motor area and left and right dorsal premotor cortex. This top-level tier was found to provide directed information to a middle tier consisting of the superior colliculi, which in turn provided directed information for the cerebellum (Fig. 3B). In addition, Lizier et al. [28] gave an in-depth discussion of the limitations in estimating neural information transfer from fMRI data due to low temporal sampling and the low pass filtering effects of the haemodynamic response function.

In sum, as in the study of functional specialization, information theoretic approaches provide model-free alternatives to commonly employed measures of functional and effective connectivity which are able to yield neurobiologically meaningful results. Future work in this line of research might address formal relationships between information theoretic and non-information theoretic connectivity estimates as well as their relative effectiveness in different experimental contexts [29].
7. Information theoretic approaches in the study of neurovascular coupling

An important question in functional brain imaging concerns the relationship between the fMRI signal and its underlying neural activity. The principal method to study neurovascular coupling is to simultaneously record electrophysiological and fMRI data and to establish some kind of statistical non-independence between different signal features of the respective modalities [30]. In the context of neurovascular coupling, information theoretic concepts have been used in a number of simulation-based studies [31–34].

For example, Nevado et al. [31] investigated the question of how neuronal population information on the level of spiking activity propagates to fMRI signal information at the voxel level. To this end, the authors took a quasi-analytical approach based on Gaussian tuning models of both single neurons and a voxel’s neuronal population and assumed a linear relationship between the sum of the spiking activity of all neurons within a voxel and its fMRI signal. In order to gain insight into questions of sensory representation, the authors used mutual information as a measure of the discriminability of a given stimulus between different voxels. The authors studied the behaviour of both neuronal population and fMRI information quantities while varying the tuning parameters of the Gaussian single neuron and neuronal population model. For example, the authors investigated whether the mutual information conveyed by the joint activity of neurons about a stimulus set can be described or predicted by measuring either fMRI signal changes evaluated at the voxel’s preferred stimulus or the fMRI mutual information. It was found that the fMRI mutual information covaries linearly with the neuronal population mutual information, provided that the tuning bandwidth is large enough with respect to the voxel’s breadth of stimulus selectivity, i.e., at high spatial resolution. In more general terms, it was demonstrated that the voxel with the strongest fMRI signal did not necessarily represent the most information. While based on low-complexity, static models of neuronal activity and haemodynamic coupling, this study nevertheless demonstrates the complexities encountered when aiming at inferring neural information encoding from functional imaging data.

As another example, Luetdke et al. [34], focussed on the question of causal rather than correlational relationships between electrophysiological and fMRI data by testing different methodologies for estimating transfer entropy on simulated data. Based on the signal statistics of an experimental simultaneous electrophysiological and fMRI data set [35], the authors generated virtual local field potentials and fMRI data that reflected the single-trial high-gamma local field potential power with Gaussian noise. The authors then proceeded to test different information bias correction methods and boot-strapping procedures on the simulated data and found that reliable and statistically significant transfer entropy estimates can be obtained even using the limited amount of data commonly afforded by empirical recordings (Fig. 3C).

These studies show that information theoretical concepts can potentially be valuable tools to infer statistical dependencies in simulation studies of neurovascular coupling and hence can help to advance our understanding of the neurobiological underpinnings of the fMRI signal. However, as yet, they have not been applied in an experimental context and their empirical success remains to be demonstrated.

8. Applications in multimodal imaging studies - an information theoretic approach to EEG-fMRI integration

The aim of simultaneous EEG-fMRI is to afford a brain imaging method combining the advantages of both modalities: high temporal resolution afforded by EEG and high spatial resolution afforded by fMRI. While some progress has been made towards its successful application, a definitive method for the analysis of simultaneous EEG-fMRI has yet to be established [36]. In the context of simultaneous EEG-fMRI, the information theoretic framework allows questions of functional specialization, functional integration and neurovascular coupling to be addressed in a unified framework.

In two studies, Ostwald et al. [37,38] investigated the applicability of information theoretic concepts to questions of multimodal integration of simultaneously acquired EEG and fMRI data. Of special interest was the question of how the insights on both the conceptual and practical information estimation obtained from the analysis of invasive neurophysiological data can or cannot be transferred to the functional neuroimaging domain. With respect to combined EEG-fMRI, the information theoretic approach allows the investigation of stimulus-coding in both domains, as well as crucial questions regarding possible synergistic or redundancy effects (from the joint response distribution), on a purely probabilistic and, hence, model-free, basis. In Ostwald et al. [37], the authors proposed the application of an information decomposition scheme [8] for the analysis of simultaneous EEG-fMRI data. Motivated by the prevalence of Gaussian models in functional neuroimaging data analysis, the study investigated the information theoretic signatures of low-complexity linear Gaussian models [39] using experimental parameters (number of possible stimuli, number of possible trials). Furthermore, for the analysis of spatially pre-selected simultaneous EEG-fMRI data, a bias correction approach combining methods developed in the analysis of neurophysiological data (PT correction [25], shuffling correction [40]) were supplemented by subtracting the remaining bias obtained in sampling Gaussian null models using the parameters of the ensuing experiment. A follow-up study [38] avoided the need to spatially preselect the input data (i.e., particular electrodes or regions of interest) by proposing a voxelwise information theoretic analysis upon projection of time and frequency domain EEG.
Fig. 4. An information theoretic argument for simultaneous EEG-fMRI recordings. This figure demonstrates the theoretical limitations of approximating a bivariate joint distribution by the product of its marginal distributions. As any information theoretic analysis is directly based on the signal’s probability distribution, the characteristics of the signal’s probabilistic dependencies are of primary importance for the successful application of information theory to empirical data. The figure demonstrates the following: if data are acquired non-simultaneously, the joint distribution of two different univariate data types (e.g., EEG and fMRI features) can only be reconstructed as the product of the respective marginal distributions, implicitly assuming the independence of the marginal random variables. However, if the joint distribution of interest is non-spherical because of dependencies between the individual random variables, its approximation by the factorized joint distribution can be very poor (see Panels B and C). This is of relevance when considering whether simultaneous or non-simultaneous EEG-fMRI should be performed for a given experiment: if one assumes that for the current paradigm EEG and fMRI features are independent (or their dependence is irrelevant), acquiring them simultaneously or separately is equivalent, because one can compute the joint distribution from the marginals. If this assumption does not hold, or EEG-fMRI covariation on fixed stimulus (conditional dependence) is of relevance, one should acquire them simultaneously. One can spin the argument even further: if one assumes that EEG and fMRI are independent, they cannot represent the same underlying neural phenomenon, as in this case they would be fully dependent. It follows that if EEG and fMRI features are independent, they must relate to different aspects of the neural response. Hence, in a generative forward modelling framework, a full picture of the neural response can only be obtained from the joint observation of EEG and fMRI features. It hence can be concluded that whether both modalities are independent or not, one should acquire them simultaneously, either to obtain the correct description of their joint distribution (in the case of independence) or to fully describe the underlying neural physiology (in the case of independence). Similar arguments could be made for the importance of assessing the topographical covariation of fMRI signals in multi-voxel pattern analyses. (A) The leftmost panel depicts the equi-probability contours of a bivariate Gaussian distribution with spherical covariance. For illustrative purposes, the distribution models the joint distribution of an EEG and an fMRI feature. The middle panels depict the respective marginal distributions of the EEG and fMRI feature by integrating over the complementary modality. Finally, the rightmost panel depicts the reconstructed joint distribution based on the product of the marginal distributions. For the case of independent marginal variables, this reconstruction is exact. (B) As in (A), but with a Gaussian mixture model representing an interesting case of EEG-fMRI dependencies: assuming that the two modes of the joint distribution are generated by different experimental contexts (e.g., different stimuli or cognitive processes), the joint distribution indicates that the experimental factor not only affects the signal’s activity dependence but also the conditional dependence (i.e., neurovascular coupling). It should be noted that this is a hypothetical example, and effects like this have rarely been demonstrated in the EEG-fMRI literature. Again, inferences based on the reconstructed joint distribution would be strongly misleading. It is within conceivable scenarios like this one that simultaneous EEG-fMRI could attain its full potential.
data into three-dimensional space based on the low resolution brain electromagnetic tomography (LORETA) formalism [41] (Fig. 3D).

On the neurophysiological level, the studies demonstrated the sensitivity of the information theoretic framework to visual contrast level encoding in both EEG and fMRI features, while questions of activity dependence, noise dependence and synergy remain largely open, due to the complexities of estimating the respective quantities.

An important aspect for future work would hence be to adapt schemes developed for the analysis of invasive neurophysiological more closely to the scientific questions of importance to simultaneous EEG-fMRI. In particular, the question of synergistic effects in simultaneous EEG-fMRI data is of primary importance for the successful integration of both modalities: synergistic effects between the two modalities would argue for a differential selectivity of EEG and fMRI features to different stimulus features and, hence, for the representation of different neurobiological substrates underlying each imaging modality. Such an observation would be interesting, because it would allow a more comprehensive characterization of neuronal processes from combined EEG-fMRI recordings than that based on the assumption that EEG and fMRI reflect the same underlying neuronal process. If the two modalities are simply different ways of measuring the same neural activity, the observation of each modality in isolation would be perfectly valid and the need for simultaneous recordings would be considerably diminished (however, see Fig. 4 for an information theoretic argument in favour of the simultaneous acquisition of EEG and fMRI data).

The information theoretic approach provides a clear framework to answer questions of this type which may be fundamental to the future exploitation of EEG-fMRI recordings.

9. Conclusion

In this review, we have discussed an information theoretic approach to the analysis of functional brain imaging data, which has its roots in the analysis of invasive neurophysiological recordings and capitalizes on the use of mutual information to describe the non-independence of experimental variables of interest. This review has demonstrated how the information theoretic framework can be used creatively in the analysis of functional brain imaging data and has hinted at some future developments for its successful use in the advancement of our understanding of brain function.

Acknowledgments

This work was supported by grant number EP/F023057/1 from the UK Engineering and Physical Sciences Research Council (EPSRC) and by the Dr Hadwen Trust.

References


