version 24 Jan 2024

Virtual electrode or virtual scalpel? A cancellation-based approach to data analysis and source localization

Alain de Cheveigné
ENS / CNRS

Abstract

The "virtual electrode" is an appealing metaphor as it suggests that we can – with the right tools and skills – use macroscopic recording techniques such as EEG or MEG to place an electrode anywhere in the brain, and record from it as if it were a real, microscopic electrode. Like every good metaphor, it can sometimes lead us astray. This short paper reviews what we can, and cannot, expect of a "virtual electrode". It is suggested that a more accurate metaphor – if less appealing – is that of a "virtual scalpel" by which activity from a source within the brain can be suppressed within the macroscopic recording. With J channels, up to $J-1$ sources can be suppressed while still retaining some information about other sources. In addition, with appropriate source and forward models, it may be possible to accurately localize any source that can be accurately suppressed. This paper sets out the possibilities and limits of what linear analysis applied to multichannel data can, or cannot, achieve.

1. Introduction

The term "virtual electrode" refers to the effect of applying, to multi-channel data, a spatial filter designed to enhance a particular source position. Specifically, given a data matrix (time × channels) recorded from electrodes or sensors outside the head, and a target physical position, the method produces a time series similar – we hope – to that recorded by a real electrode at that position (Fig. 1). The term is motivated by the fact that the filter targets a particular *spatial position*, and produces a single *time series* as might a real electrode. The virtual electrode can be virtually positioned anywhere within the brain, and the associated spatial filter mimics – we hope – spatial attenuation with distance from the tip of a real electrode. There has been some empirical success, for example in recovering the position and time course of sources picked up with real electrodes in epileptic patients undergoing pre-surgical recording with stereotactic EEG (sEEG) electrodes inserted deep in the brain (e.g. Cao et al 2022, Coelli et al 2023).

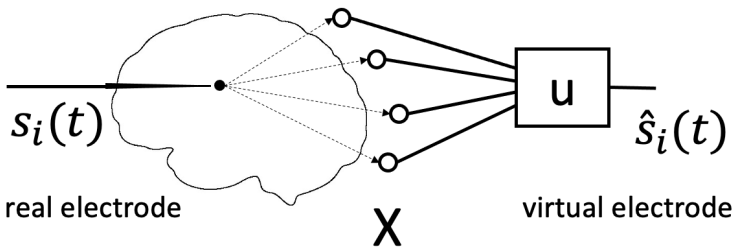

Fig. 1. A real electrode (left) is selectively sensitive to brain sources near its tip, and blind to distant sources. A virtual electrode (right) attempts to reproduce this property by applying a spatial filter to the multichannel observation matrix **X**.

The prospect of clinical applications is appealing and motivates efforts to refine the methods and extend their applicability. A hospital administrator that hesitates at the cost of installing and running a MEG system might be swayed by the perspective of using it to save time, cost and risk to patients over surgical implantation of deep electrodes. It is important for him or her (and for the patients) to know whether a virtual electrode can supplant a real electrode in this context.

A simple mathematical argument tells us that a "virtual electrode" is not a real electrode. A real electrode can be placed anywhere in the brain, potentially picking up activity different from any other location. Supposing that electrodes could be placed in each of I distinct

locations (e.g. one in each of the ~120000 cubic millimeters of the human brain), recording from them simultaneously would result in a data matrix of rank up to I. In contrast, the same number of virtual electrodes can result in a data matrix of rank no greater than J, the number of sensors or electrodes (e.g. 128 for a typical EEG system). This is because all the virtual electrode waveforms reside in the space spanned by the J electrodes they are formed from. Since $J \ll I$, these matrices are clearly different: virtual electrodes are not the same as real electrodes.

A similar argument can be made with respect to imaging (for example magnetic source imaging, MSI). At any point in time, only J measured values are available to construct the I pixels or voxels of the "image", contrary to an optical imaging technique that offers $I \gg J$ independent observations. Likewise, the time course of any pixel or voxel is a function of just J sensor signals, in contrast to the $I \gg J$ time courses – potentially independent – of an optical imaging technique.

These limitations are, of course well known, as they reflect the fundamental indeterminacy of the inverse problem (Baillet 2010). As in many other estimation tasks – and indeed perception itself (Helmholtz 1867) – additional information is required, for example in the form of an internal model constrained by the observations. However, this is problematic in an observation tool, because the limits between *observation*, *prior*, and *inference* are blurred. Is the boundary of cortical activity, apparent in the image, a reflection of the data, or of the prior? Conversely, is imperfect localization or imaging the result of basic indeterminacy, insufficient or noisy data, issues with sensor calibration or head position tracking, imperfect knowledge of anatomical properties (e.g. head conduction), or suboptimal properties of the analysis or imaging method? These factors are not always well separated in the literature, making it hard to judge between method- or equipment-related limitations (that are amenable to progress) and fundamental limitations (that are not). A goal of this paper is to cleanly identify and separate these factors.

In summary, the metaphor of virtual electrode is useful to convey some of the power of magnetic or electric source localization and imaging approaches, and foster enthusiasm for them, but it carries with it the seeds of misunderstanding their limits, and whether or those limits can be overcome by better skill, algorithms or hardware, or not. This paper explores a different metaphor, less enticing but possibly more accurate, that of a “virtual scalpel”. While

it is not possible, in general, to focus on a source and extract it from within the recorded data, it may be possible to *silence* that source, and by doing so, in some situations, locate it with precision.

2. Concepts and methods

I first describe the theoretical background, with the slightly unusual perspective of selectively *suppressing* a source, as opposed to isolating or enhancing it (terms "suppress", "cancel", "silence" will be used interchangeably in this sense). I then present a "toy world" with a simplified planar source geometry and monopolar field patterns to allow easy visualization using 2-D plots (as opposed to the 5 dimensions required to specify three spatial coordinates and two orientation parameters). Finally, I attempt to draw conclusions applicable to sources within a real brain as measured for example using EEG or MEG.

3.1. Data model

Recorded data take the form of a matrix $\mathbf{X}$ of size $T \times J$ where T is the number of samples and J the number of channels (electrodes or sensors). The recording is assumed to reflect I sources within the brain via a linear mixing process:

$$\mathbf{X} = \mathbf{SM}$$

where $\mathbf{S}$ of size $T \times I$ represents the activity of the sources and $\mathbf{M}$ is a mixing matrix of size $I \times J$ (Fig. 2A). The matrix $\mathbf{X}$ is observable, $\mathbf{S}$ and $\mathbf{M}$ can only be inferred. There are many more sources in the brain than recording channels: $I \gg J$.

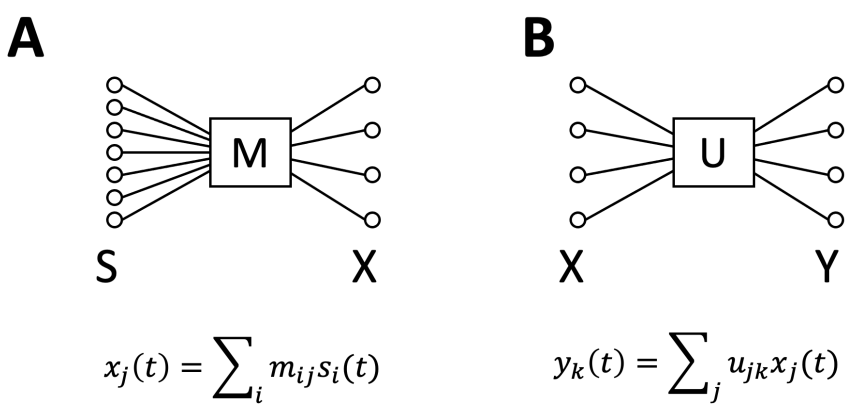

Fig. 2. A: Data model. The signal of each of the J observed channels is the weighted sum of I brain source signals. B: Analysis model. The signal of each of the K components is the weighted sum of J observed channels.

3.2. Analysis model

Typically, we are interested in (a) one or more columns of **S**, representing the *time series* of activity of brain sources of interest, and (b) the corresponding columns of **M** insofar as they might allow to infer the *spatial positions* of those sources. Neither **S** nor **M** are observable, but it might be possible to infer them from the data, at least in part.

Since the sources were mixed linearly, it makes sense to try to recover them using linear methods. Typically, an analysis matrix **U** of size $J \times K$ is applied to the data to yield a matrix **Y** of size $T \times K$ of transformed data (Fig. 2B). Each of the K columns of **Y** (a time series) is referred to here as a *component*. K can be smaller or larger than J, however if it is larger the components are linearly dependent and thus correlated: the rank of **Y** cannot be greater than that of **X**. At the other extreme, if $K = 1$, the single-column matrix **U** is a *spatial filter*. The matrix **U** might be predefined, as in re-referencing, Laplacian, or current source density (CSD) transforms, or instead derived from the data, as in principal component analysis (PCA), independent component analysis (ICA), or other blind source separation (BSS) methods.

Ideally, we would like for the analysis matrix to be an "unmixing matrix" that recovers the time course of each brain source. In that case, **MU** would be the identity matrix, and each column of **Y** would map to a source. Unfortunately, brain sources number in billions, whereas sensor channels range from tens to thousands, so recovering **S** is impossible. We must instead settle for the less ambitious objective of finding components that are somehow informative. For example, a column of **Y** might approximate the activity of a particular source, or compound activity with a functional meaning such as an event-related potential (ERP).

3.3. Vector spaces

As mixing and unmixing are linear, it is helpful to see them as operating within a *vector space* of signals: concepts such as subspace, span, dimensionality, linear dependence and orthogonality allow for useful insight and short-cuts. The signal $x_j(t)$ picked up by an electrode or sensor belongs to the space *spanned* by the brain source signals $s_i(t)$. That space is of *dimension I* at most, although it might be smaller if some of the $s_i(t)$ were *linearly dependent*. The J columns of the observation **X** then span a *subspace* of that space, of dimension of at most J. The mixing matrix **M** projects from the source space onto this subspace,

and thus a given observation **X** can reflect many different patterns of brain activity (Fig. 3), a property that is important later on.

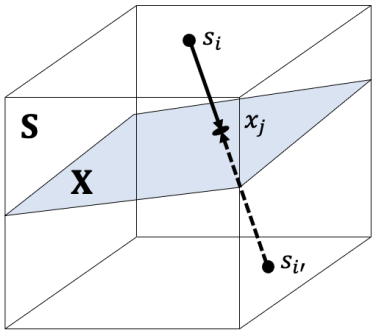

Fig. 3. The space spanned by the columns of the observation matrix **X** is a subspace of the space spanned by the brain signals **S**. Brain source signals are projected onto this subspace, and thus multiple sources may map onto the same observation.

From the observation **X** is linearly derived the component matrix **Y**, which thus spans a subspace of the observation space spanned by **X**. Some analysis methods (e.g. PCA) result in columns of **Y** that are mutually uncorrelated. In that case, the subspace spanned by any subset of columns is *orthogonal* to the subspace spanned by the other columns: each vector of one subspace is uncorrelated with every vector of the other subspace.

3.4. Spatial filter and virtual electrode

If $K = 1$, the analysis matrix consists of a vector **u** of weights (spatial filter). Here, the word "spatial" refers to the J-dimensional signal space, although it is sometimes also understood in the sense of selectivity in the physical space of brain sources (the term "virtual electrode" evokes selectivity in that sense). Spatial filters too form a vector space, since a linear combination of spatial filters is a spatial filter. The dimension of that space is J, because each filter is defined by J coefficients.

Spatial filters have several nice properties:

1) For every source $\mathbf{s_i}$ there exists a spatial filter **u** that can suppress it. A trivial example is a filter with all coefficients zero except for two channels j and j' such that $u_j/u_{j'} = -\,m_{ij}/\,m_{ij'}$. I call this a *null filter* for that source.

2) A null filter projects from the I-dimensional source space $\mathbf{S}$ to the one-dimension filter output, and thus other sources must map to zero (Fig. 3). I call these sources the *null set* of the filter (see Figs. 4C and following for illustrations).
3) If $J > 2$, there exist additional filters also that suppress $\mathbf{s}_i$. Together, all of those filters form a subspace of dimension $J - 1$ of the space of filters over J channels.
4) Extending 1), for every set of $J - 1$ sources, a filter can be found that cancels those sources simultaneously. In other words, with J observation channels one can suppress at least $J - 1$ sources (see Fig 6 for examples).

Herein lies the power of a spatial filter: the ability to *suppress* or silence irrelevant or obnoxious sources. The output of a spatial filter is a single time series, invariant to the activity of the suppressed sources, and that reflects the compound activity of all the other sources, those that were not silenced.

A real electrode is sensitive to sources near its tip, and insensitive to distant sources (Fig. 1, left). To reproduce that property, a spatial filter (Fig. 1 right) must put a null on each of those distant sources. In other words, most coefficients of $\mathbf{Mu}$ must be zero. As they number in billions (I), whereas the free parameters of a filter are fewer (J), this is a hopeless task. Herein is the basic limitation of the concept of "virtual electrode": a virtual electrode, contrary to a real electrode, cannot select one source and ignore all others. It can only ignore $J - 1$ of them.

3.5. Data-driven analysis

So far, we discussed analysis matrices $\mathbf{U}$, or spatial filters $\mathbf{u}$, arguing that they must exist but not saying how they are found. The analysis matrix $\mathbf{U}$ is sometimes predetermined, but more often it is calculated based on the data using an analysis method such as PCA, ICA, or others.

Among data-driven methods, PCA stands out for three useful properties. First, the columns of the component matrix $\mathbf{Y}$ are mutually uncorrelated. Second, their variance adds up to the variance of $\mathbf{X}$. Third, the first principal component (PC) packs the largest possible amount of variance, the second the largest amount of variance in the subspace orthogonal to the first, and so-on. An interesting consequence of this last property is that the *last* principal component includes the *least* possible amount of variance (zero if the data are rank-deficient). The last column of $\mathbf{U}$ thus defines the spatial filter that best suppresses all sources present. If $J - 1$

sources were active, the last column of the PCA matrix would map them to zero, thus defining a spatial filter that cancels them all.

Other methods offer different properties. ICA attempts to ensure that components are *statistically independent*, in addition to uncorrelated. Canonical correlation analysis (CCA) seeks linear combinations of columns of **X** that are *maximally correlated* with another dataset $\mathbf{X}'$. Joint decorrelation (JD) seeks to maximize the *ratio of variance* between the data and a filtered version of the data (see de Cheveigné and Parra 2005 and references therein). These methods share with PCA the property that columns of **Y** are mutually uncorrelated.

Classic studies focus on the ability of a spatial filter to *select* an individual source or group of sources. Here, we focus instead on its ability to *exclude* one or more sources (up to $J - 1$). The filters are the same, the difference is in their interpretation.

3.6. Existence vs identifiability

The analysis matrix **U** obtained by a data-driven method might fail to produce the desired outcome. It is useful to keep in mind that this can happen in at least two ways: (a) *the desired solution exists but we fail to find it*, for example due to overfitting or geometrical inaccuracy, or (b) *the desired solution does not exist*, as a result of more fundamental constraints. In the former case, a new technique might overcome the limitation; in the latter case, not.

3.7. Localization and imaging

The discussion so far made no reference to the physics or geometry of the problem. Linear analysis methods are agnostic to those aspects, and in return they have nothing to say about the position of a source.

To localize a source, we need to introduce a *source model* that constrains (or prioritizes) possible source positions and/or shapes within the brain, and a *forward model* that specifies the source-to-sensor gain coefficients for each allowable source position. The source model is required because with $J \ll I$ sensors the source localization problem is highly unconstrained. Common approaches to source localization include *dipole fitting*, *scanning* with a beamformer, or *inversion* of a source-to-sensor matrix under particular constraints (Lütkenhöner 1998, Baillet 2001, Hauk et al 2022). Here, I suggest an additional approach.

The principle is simple. As mentioned earlier, each filter is associated with a *null set* that consists of all the sources that map to zero via that filter. Thus, if a source is cancelled by a filter, its position is constrained to that set, and if it is cancelled by multiple filters, to the *intersection* of their null sets. Concretely:

1) Using the source model, define the set of possible sources. For example, it might include radial dipole sources constrained to the surface of the cortex, or extended sources, or bilateral pairs, etc.
2) Using the forward model, create a dictionary of source-to-sensor gains. It can be represented by a $I \times J$ matrix $\mathbf{D}$ exhaustively over a high-resolution grid, or via an interpolation procedure to save space or computation.
3) Using a data-driven analysis method, find one or more spatial filters that zero the source of interest. For example, a method such as ICA might isolate that source within column $\mathbf{u}_j$ of the analysis matrix, in which case the $J-1$ other columns $\mathbf{u}_{j'}, j' \neq j$ constitute null filters for that source.
4) For each null filter, form the product $\mathbf{D}\mathbf{u}_{j'}$ and search for indices (source locations) for which it is zero. These define the null set of that filter within source space.
5) The intersection of null sets of all null filters further constrains the location of the source, possibly to a single point (within the parameter space of the source model).

The accuracy of localization depends on (a) the completeness of the source model, (b) the accuracy of the forward model, and (c) the accuracy of the data-driven analysis method. If all three are granted, there is no limit to accuracy (no equivalent of a "point-spread function", e.g. Hauk et al 2022). If enough filters are available the location may be constrained to a point, otherwise it is extended or ambiguous.

This process can be extended to derive a visual rendering of inferred brain activity. However, the rendered image depends critically on the source model *and on the convention chosen to visualize points within the parameter space of that model*. If the source model includes point sources, and these are rendered as points within the image, the image will suggest punctuate activity. If the source model allows for (or imposes) extended sources, and these are rendered according to their extent, the image will suggest extended activity. The data are the same in each case, and thus the image does not support one aspect or the other (punctuate or extended).

A particular choice of source or rendering model might make the image appear to be anatomically realistic, however the image cannot be used as *experimental support* for all the anatomical details that it shows.

4. Results

4.1. A toy world

The previous concepts are abstract. In particular it may be hard to grasp the relations between *signal* space, *parameter space*, and *physical space*. This section introduces a setup with simplified geometry to help clarify geometric versus algebraic aspects. Sources and sensors are within a plane, and source-to-sensor propagation depends only on the distance between source and sensor. The geometry is thus fully specified by the coordinates of sources and sensors within this 2-D space.

Specifically, a set of sensors is positioned along the margin of a disc (Fig. 4 A, blue dots), while sources are located within that disk. The activity of each source is assumed to radiate within the plane with an amplitude that decreases in an isotropic manner as $1/r^2$ (illustrated by pink shading for a single source, green cross). Source-to-sensor gain is thus inversely proportional to the square of the source-to-sensor distance. If a sensor could be placed at the position of the green cross (Fig, 4 B), its gain pattern (gray shading) would be selective to that particular source. Unfortunately, sensors are constrained to the periphery, and their gain patterns are unhelpful for selecting sources in positions other than near a sensor (Fig. 4 C).

However, with multiple sensors (here 5), one might hope to synthesize a gain pattern that mimics Fig. 4 B, i.e. create a "virtual sensor". We shall see that that goal is hard to reach, but a different goal is easy: any particular source can be accurately *cancelled*. The weighted sum of two sensors $y(t) = u_j x_j(t) + u_{j'} x_{j'}(t)$ forms a spatial filter with a position-dependent gain, as illustrated in Fig. 4 D. Here, the coefficients were chosen so that the gain is zero for a source at the position of the green cross. Interestingly, the gain seems to be zero along a line, implying that sources along that line would also be suppressed. This is referred to as the "null set" of the filter, while the filter itself is referred to as a "null filter" for that source.

This illustrates three of the properties mentioned earlier. First, for any source, there exists a spatial filter that can suppress it (a null filter for that source). Second, that filter suppresses

multiple sources (the null set). Third, that filter is sensitive to sources situated in most other locations. From the curvilinear aspect of the null set in Fig. 4 D, it is tempting to assume that it constitutes a manifold of the 2-D space of locations. However, some examples seem inconsistent with a manifold (e.g. Fig. 6 D), and in any event that property is not required here.

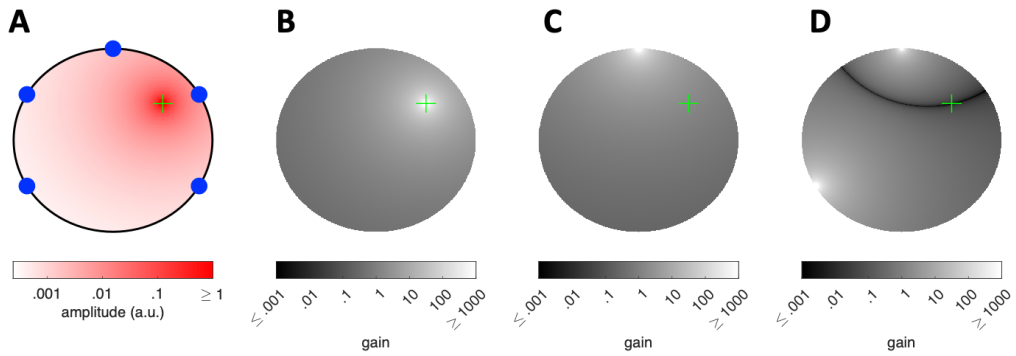

Fig. 4. A: Simple setup consisting of an array of sensors (blue dots) placed at the periphery of a disc. A source within the disc (green cross) radiates within the plane with an amplitude that decreases with distance as $1/r^2$ (pink color scale). B: A sensor placed at the cross would have a gain pattern spatially selective for this source (gray scale). C: Unfortunately, sensors are constrained to the periphery, with unhelpful gain patterns. D: Combining those sensors linearly can produce a gain pattern that accurately *rejects* the source. The dark line indicates other locations of zero gain (the "null set" for this filter).

Figure 5 shows gain maps for four additional null filters relative to the source situated at the green cross.

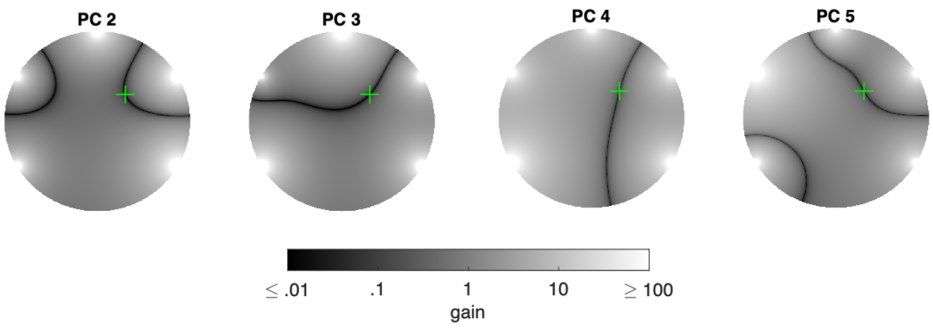

Fig. 5. Gain maps for spatial filters derived from columns 2 to 5 of the transform matrix of PCA applied to data from 5 sensors in response to a single source located at the green cross. All four filters are null filters for that source: the source is located within the null set of each filter.

4.2 PCA

The filters in Fig. 5 were obtained by applying PCA to a matrix obtained by simulating a Gaussian noise source and applying the forward model to obtain a simulated 5-sensor data matrix. For a single source, that matrix is necessarily of rank one, and thus the last four columns of the PCA matrix map the data to zero, i.e. they are null filters for that source.

This illustrates the property, mentioned earlier, that multiple spatial filters exist that can suppress a single source, as long as $J > 2$, which is the case here. Not only do they *exist*, but they *can be found* using PCA if only one source is active: columns 2 to 5 of the PCA matrix map to zero, i.e. those columns are null filters relative to that source. Any linear combination of null filters is also a null filter and the columns are mutually orthogonal (a property of the PCA transform), and thus those filters span a space of filters of dimension 4, subspace of the space of 5-channel spatial filters. To each of these 4 filters is associated a null set, each with a different shape, that all pass by the source (Fig. 5).

4.3. Multiple sources

What if there is more than one source? The PCA method is still applicable as long as $I < J$: the matrix of sensor data is of reduced rank and thus one or more columns of the PCA matrix are null filters for all the sources, simultaneously. Here, for two sources there is a subspace spanned by three filters (Fig. 6 A), for three sources by two filters (Fig. 6 B), and for four sources by a single null filter, unique to a multiplicative constant (Fig. 6 C).

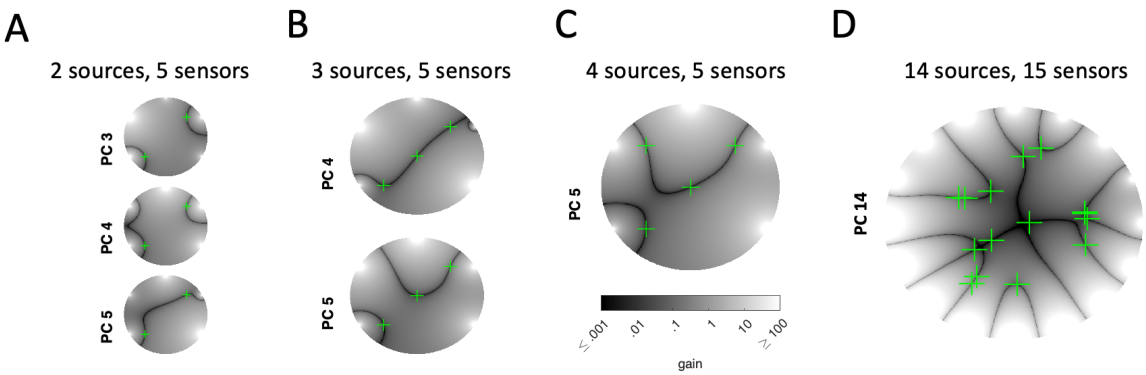

Fig. 6. Gain maps for null filters for 2 sources (A), 3 sources (B) and 4 sources (C), assuming 5 sensors. D: same for 14 randomly-placed sources and 15 sensors.

Figure 6 D shows another example with 15 sensors and 14 randomly placed sources. The null set in this example is extensive and intricately branched, but still sparse within the source space: despite zeroing those 14 sources (and more), this filter is sensitive to most sources within the source space.

In summary, a filter *exists* that completely silences up to $I < J - 1$ unwanted sources while remaining sensitive to activity elsewhere in source space except for a limited subset (null space). If $I < J - 1$, there are several such filters that all belong to a subspace of filters of dimension $(J - I)$. Those filters can be *found* using PCA, as long as the $J - I$ sources are the only ones active and their time courses are not linearly dependent. Approaches to estimate the filters when $I \geq J$ are discussed later.

4.4 Localization

Considering the example of Fig. 4 D, suppose that we do not know the position of the source, but we know that the 2-channel filter suppresses it. Given the source and forward models, we can conclude that the source lies somewhere along the dark line. In the same way, each of the filters in Fig. 5 is a null filter for the source at the green cross, and thus we know that that source is within the null set of each of those four filters.

In our toy world (Fig. 4 A), the "source model" is the disc, the "forward model" is embodied by the sensor positions and the $1/r^2$ law that governs source-to-sensor gain, and a null set is a subset of the disc. In a more realistic situation, sources might have a 3D position and orientation, with anatomical constraints, or a "source" might be bilaterally distributed or spatially extended. The null set would then be some subset of the parameter space of that source model.

In Fig. 5, each filter has a different null set, and they all intersect (by construction) at the source that they are designed to suppress. Their intersection is a single point, as obvious by taking the max over gain maps (Fig. 7 A). The precision of the location is limited only by numerical or physical precision (here, the 500×500 resolution of the search grid). There is no equivalent of "point spread", as invoked to quantify localization uncertainty in MSI (e.g. Hauk et al 2022).

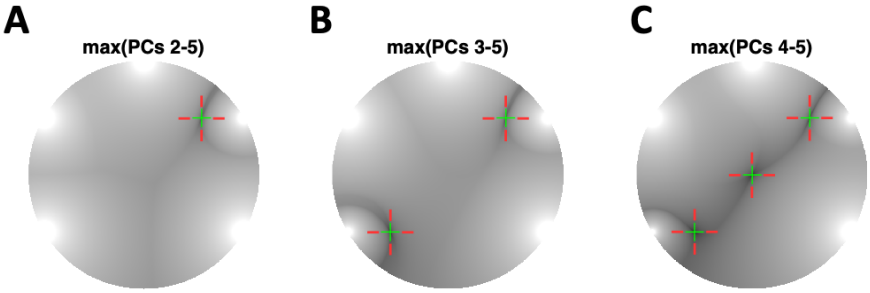

Fig. 7. A: Maximum of gain maps of four null filters relative to one source (individual maps plotted in Fig. 5). B: Maximum of gain maps of three null filters relative to two sources (individual maps plotted in Fig. 6 A). C: Maximum of gain maps of two null filters relative to three sources (individual maps were plotted in Fig. 6B). Green crosses indicate the ground truth of the source positions, red crosses indicate the estimated positions (local minima of the maximum of gain patterns).

For one source and J sensors, there exist $J - 1$ null filters, each with a distinct null set. This implies a very strong constraint on the location of the source *if those null filters can be found*. In Fig. 5, all four filters could be found using PCA because only one source was active. In the configurations of Fig. 6 A and B, respectively two and three sources were active, and hence only three and two filters could be found using PCA. The corresponding max gain plots are shown in Figs. 7 B and C. Each shows multiple zeros (2 and 3 respectively), with positions accurate but ambiguous as to which source lies where. The fewer filters, the harder it is to locate the sources, and in the limit case of four active sources, only one null filter can be found, defining one null set, and we can say only that the sources are somewhere within that single, spatially-extended null set (Fig. 6 C).

4.5. Source Separation and localization

The PCA-based localization method less effective as the number of sources approaches J, and inapplicable if $I \geq J$. However, a number of methods can be substituted for PCA that are less constrained. For example, ICA aims to separate sources that are statistically independent according to an empirical measure of independence (that depends on the algorithm). If the algorithm succeeds in isolating a source of interest within column j of the analysis matrix $\mathbf{U}$, then *all other columns are null filters for that source*. Similar algorithms such as JD can be used in the same fashion.

To illustrate this idea, JD (a.k.a. DSS or CSP) was applied to simulated data obtained by mixing five sources into five sensors. The source signals were derived from 64-channel EEG data recorded in response to repeated presentations of an auditory stimulus (Fuglsang et al 2020). Data of one subject were processed using the JD algorithm to isolate the most reliably repeatable component, and this component was used here as the time series of one source. Four other sources were obtained from randomly-selected individual EEG channels of the same recording, after regressing out the repeatable component. The sensor data, obtained from the source data by applying the toy-world forward model to these sources, are thus similar to real EEG but with precisely-known source locations. These sensor data were processed using the JD algorithm to find a transform matrix that isolates, within its first column, the most repeatable source. Of interest here are the *other* columns, those that are null filters for that source.

Figure 8 A shows the gain maps for the four null filters. The green cross is the target source (repeatable) and the blue crosses are the other sources. The null sets all pass via the target source, as obvious from a plot of the maximum over gain maps (Fig. 8 B). The minimum within this plot (red cross) coincides with the location of the source (green cross). Figure 8 C shows a similar plot for 16 regularly-spaced sensors and 16 randomly-placed sources (derived from the same EEG data). Again, the estimated location (red cross) coincides with the ground truth (green cross). It should be mentioned that localization fails for a small proportion of simulations (not shown), presumably because certain randomly-drawn configurations make the problem undetermined.

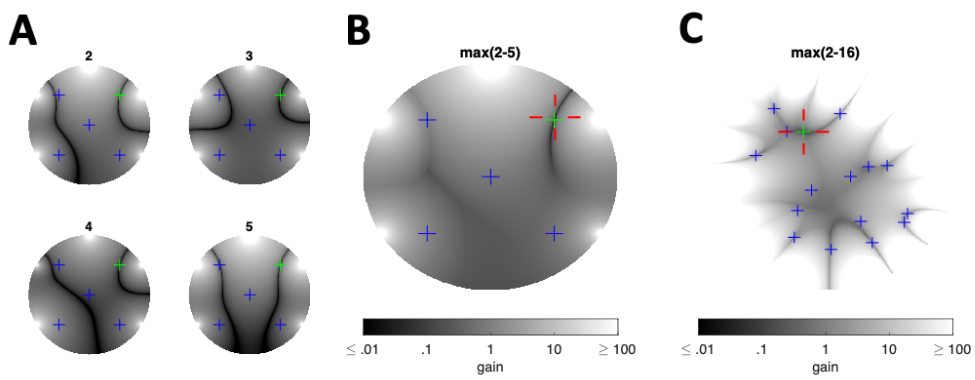

Fig. 8. A: gain maps for columns 2-5 of the transform matrix obtained by applying the JD algorithm to 5-channel sensor data obtained by mixing five sources, one of which is repeatable over trials (green), the other not (blue) (see text). B: max over gain plots in A. The red cross indicates the

estimated location of the targe source (green cross). C: same, for 16 sensors and 16 randomly-placed sources.

If the source-separation algorithm successfully isolates a source of interest, there are $J-1$ columns of the analysis matrix to serve as null filters for that source, which is strong constraint on the source location. The accuracy of localization then depends mainly on the accuracy of the source and forward models. In addition to repeatability, other criteria can be used to bias JD, for example spectral content (e.g. oscillations) or non-stationarity (e.g. event-related desynchronization) (see review in de Cheveigné and Parra 2014). Other algorithms can also be used, such as CCA to isolate a brain source that correlates with a stimulus. ICA itself comes in many flavors, exploiting various measures of statistical independence (Cardoso 2001).

All these algorithms (including ICA and JD) depend implicitly on the assumption $I \leq J$. Strictly speaking, this assumption is unrealistic, if only because sensor noise alone contributes J sources, to say nothing of the billions of brain sources. If $I > J$, the assumption may remain valid as an approximation, for example if applied to time intervals where most sources are quiescent. Ultimately, this is an empirical question.

4.5. Non-punctuate sources

So far, sources were assumed to be points within a 2D disc, but the methodology can be extended to more complex configurations. As an illustration, assume that sources are either unilateral or bilateral (for example lateralized vs bilateral cortical activity). To handle this, the source space can be augmented to include symmetric bilateral sources (both sides with the same activity) as well as unilateral. The null space of a filter is then defined within this augmented space. Figure 9 shows the maximum over gain maps for four null filters for a source that is either bilateral (A) or unilateral (B). For a bilateral source, the unilateral source space contains no null space and thus no clear mimina (A, left) but the bilateral source space shows a clear minimum (red cross). For a unilateral source, the unilateral source space shows a clear minimum (red cross) while the bilateral source space contains none (B, right). Thus, it is possible to determine (a) whether the source is unilateral or bilateral, and (b) its accurate position.

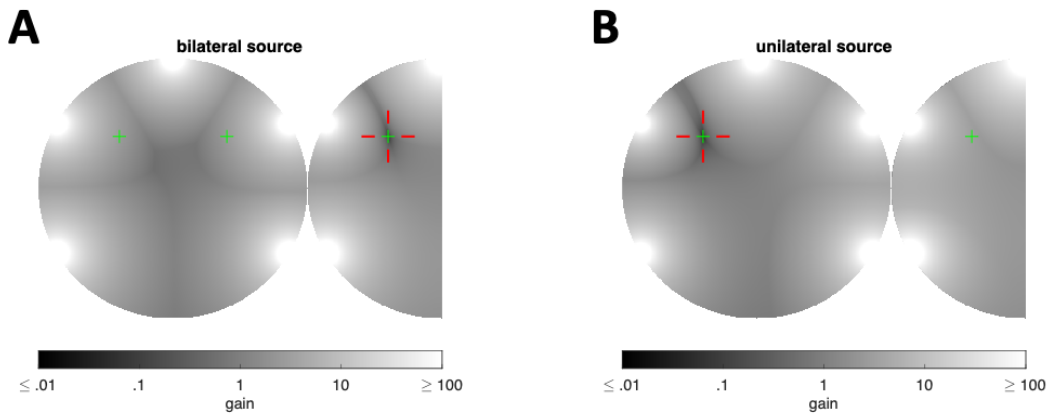

Fig. 9. A: Maximum over gain maps for four null filters for a bilateral source. The left part represents the source parameter space for unilateral sources, the right part (half disc) the space for bilateral sources. B: Same, for a unilateral source. The green crosses indicate the potential source locations, the red crosses indicate the estimated positions.

Application to more realistic configurations (positions in 3D space, dipole orientations, extended sources, anatomical constraints, realistic head conduction models, etc.) is beyond the scope of this paper.

4.6. Beamformer

It is interesting to contrast the present approach with a standard method for data analysis and localization, the linear constrained minimum variance (LCMV) beamformer. The LCMV algorithm produces a spatial filter that (a) passes activity from one specific location, and (b) minimizes output variance, thus suppressing activity from sources located elsewhere. These are properties expected of a virtual electrode, and indeed the LCMV beamformer is a popular method to implement one. The desired spatial filter $\mathbf{u}$ is obtained as:

$$\mathbf{u} = (\mathbf{f}^{\intercal}\mathbf{C}^{-1}\mathbf{f})^{-1}\mathbf{f}\mathbf{C}^{-1}$$

where $\mathbf{C}$ is the covariance matrix of the data and $\mathbf{f}$ is the vector of source-to-sensor weights for that position according to the forward model. This solution requires taking the inverse of the covariance matrix, and thus is applicable only if the observation matrix is of full rank, i.e. $I \geq J$. Furthermore, we saw earlier that the goal of suppressing all sources other than the target is possible only if they number at most $J - 1$. Thus, the solution can be accurate (perfect interference rejection) only if $I = J$.

Figure 10 gives an example of applying the algorithm to a 5-source, 5-sensor configuration, with the beamformer targeting the upper right source (dotted circle). The resulting spatial filter perfectly suppresses all other sources while preserving the target source, as expected. The gain map is actually identical to that plotted in Fig. 6 C, consistent with the earlier argument that a filter suppressing 4 sources within data from 5 sensors is unique. However, the beamformer solution is somewhat fragile. The solution illustrated in Fig. 10 A was obtained with data from sources that were *perfectly decorrelated* (this was ensured by orthogonalizing the matrix generated by the Gaussian noise process used to simulate source activity). Omitting that step yields the solution illustrated in Fig. 10 B. In the presence of mild inter-source correlation (here $|\rho| \leq 0.05$), the beamformer may fail to suppress all unwanted sources.

Even in the case where beamforming is successful, the properties of the solution are not consistent with the idea of a virtual electrode. Based on dimensionality considerations (Fig. 3), it was argued that an extended set of sources is equivalent from the point of view of a spatial filter. Thus, the solution cannot be perfectly spatially selective: a virtual electrode pointed to one location will pick up activity from other locations. Indeed, Fig. 10 A shows that the gain of an "electrode" "pointed at" the red source is non-zero for most other locations within the disc. This is certainly very different from the gain pattern of a real electrode at that position (compare Figs. 10 A and 4 B).

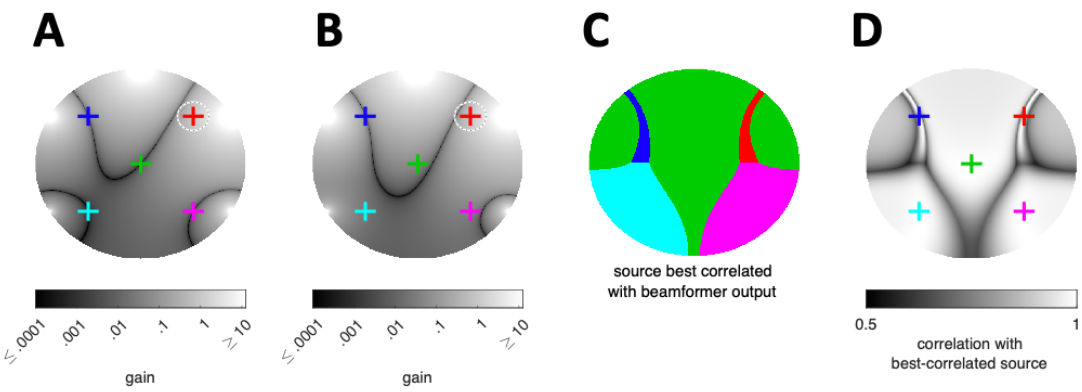

Fig. 10. A: gain map of a beamformer tuned to target one source among 5 active sources (crosses). The targeted source is indicated by a circle. B: same, but the source time series of are slightly correlated between sources ($|\rho| \leq 0.05$), resulting in an imperfect solution. C: identity of the source that is best correlated with the beamformer output as the beamformer is scanned across the disc. Color codes are the same as for the crosses D: correlation between the beamformer output and the best-correlated source.

The LCMV beamformer has been proposed as a means to localize a source, by scanning the source space and monitoring the filter output (e.g. for spectral content, or ERP amplitude). Figure 10 C and D show the result of such a scan, evaluated by measuring the correlation of the output with each of the sources. Panel C shows the identity of the best- correlated source as a function of target position: the source space is partitioned into zones which each include, as expected, the appropriate source. This provides a rough basis for localization, which we might hope to refine the locus by searching for a *peak* in the value of the correlation. However, panel D shows that such peaks are extremely wide. The two topmost sources are constrained each to a relatively narrow "horn", but for the others the patterns are wider. One might be tempted to view such diffuse patterns as consistent with anatomical spread, but our knowledge of the ground truth here tells us that that view is unwise.

5. Discussion

This paper considers data analysis from an unusual angle, regarding spatial filtering as a means to *suppress*, rather than *select*, specific sources of brain activity. The mathematics are the same, the difference is mainly in the interpretation. I have tried to clearly separate *algebraic* aspects (existence of filters, their vector-space structure and dimensionality), *data-driven estimation* (PCA, ICA, etc.), *source models* and *forward models* relevant for source localization, and *rendering models* relevant for imaging. Three main conclusions can be drawn.

5.1. The term "virtual electrode" should be retired

That the term is an approximation, or metaphor, is well known. Less widely appreciated is just how far it is from being accurate. It is clear from theoretical arguments (Fig. 3) that a spatial filter is much more "promiscuous" than a real electrode, and this is confirmed by simulations (e.g. Fig. 10). A spatial filter, however it is called, cannot approximate the spatial selectivity of a real electrode. The risk, of course, is that layman or specialist alike are misled into thinking that it can. Given that risk, the term is best retired.

5.2 Linear analysis is useful to *suppress* (rather than enhance) sources

Related to the previous point, a spatial filter cannot selectively enhance a particular source, except as a side-effect of suppressing a limited number of competitors. That being the case, it is more straightforward to directly reason in terms of suppression. A spatial filter can suppress

up to $J-1$ competing sources while remaining sensitive to other sources, hopefully including those of interest to the investigator. There is unfortunately no guarantee that sources of interest won't fall within the null set of the filter, but if enough sensors are available, there may be multiple filters that can suppress the competing sources, each with a different null set, increasing the likelihood that at least one of them will preserve the sources of interest. This is an argument in favor of high-density recording devices.

The required spatial filter(s) can be predetermined (rarely) or derived from the data using algorithms such as PCA, ICA, CCA, JD, and others. Data-driven analysis can fail in either of two ways. It may fail because the required filters do not exist (desired and undesired sources are not linearly separable), or because they cannot be accurately estimated, for example because of overfitting. The latter failure can take two forms, of unequal gravity: the target activity may remain unobserved (the algorithm failed to identify the filter that would reveal it) or a spurious observation may result from fitting to a quirk in the data. Cross-validation methods may help detect overfitting, and regularization prevent it from happening.

Suppose that we can find a transform $\mathbf{U}$ that produces components that are good proxies of sources, for example using ICA or joint diagonalization (JD). One or several sources can then be silenced by multiplying those components by zero and applying the inverse of $\mathbf{U}$. A "denoising matrix" can thus be defined as

$$\mathbf{D} = \mathbf{U}\mathbf{E}\mathbf{U}^{-1}$$

where $\mathbf{U}^{-1}$ is the inverse (or pseudoinverse) of the transform matrix $\mathbf{U}$, and $\mathbf{E}$ is a diagonal matrix with ones for components to retain and zeros for components to remove. Applying this to the data yields a "clean" data matrix:

$$\widehat{\mathbf{X}} = \mathbf{X}\mathbf{D}.$$

The matrix $\widehat{\mathbf{X}}$ has the same dimensions as the original data and can be substituted for it in further analyses, with two caveats. First $\widehat{\mathbf{X}}$ is not of full rank. Second, the forward model may need to be adjusted for the non-uniform gain due to the denoising matrix.

5.3 Accurate localization is possible given accurate source and forward models

Our ability to localize and image sources are limited by the fundamental indeterminacy of the inverse problem (Baillet 2010). Methods differ according to how they deal with those limits. Here, they are dealt with in a modular fashion, according to the flow-chart of Fig. 11.

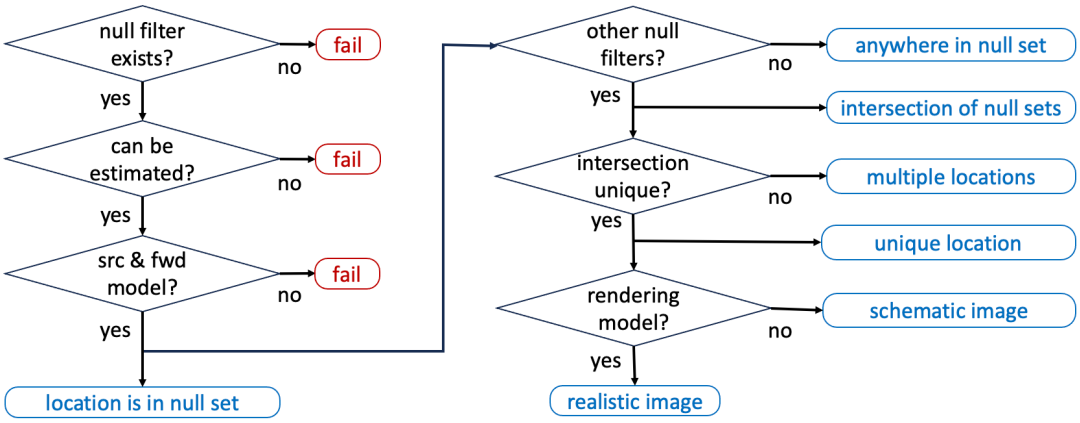

Fig. 11. Flow chart for source localization and imaging. Basic conditions (left) are that at least one null filter exists and can be estimated, and that a source model and a forward model are available. With one null filter, the source location is constrained within its null set. With several null filters, it is constrained within their intersection. If the intersection is not unique (e.g. Fig. 7 B and C), the location is ambiguous. If it is unique, the location is known within the space of source parameters. If a realistic rendering model is available (e.g. smoothing, etc.) the image may appear realistic (e.g. distribution over a cortical area), otherwise it remains schematic.

An appeal of this approach is that it delineates the requirements of each step (and their failure modes), clearly separating issues of *existence* of linear solutions, *estimation* of those solutions, *physical and anatomical constraints* on sources, *accuracy* of the forward model, and *cosmetic* tricks to improve the realism of images. This is in contrast to papers on these topics in which these issues are confounded. Source locations within the source parameter space are found by search, which can be interpreted in terms of Bayesian estimation, the source model serving as prior (probability one within allowable space, zero elsewhere) and the forward model + filter as evidence (probability one if the output is zero, zero otherwise).

5.4. Future directions

The reader might remain unsatisfied by the limited space devoted to some important issues. This paper should be seen as setting the stage for future research to address them, full resolution being outside its scope.

5.4.1 Application to real EEG and MEG

An obvious future direction is application to real data. The "toy world" allowed for clear, unambiguous visualization within a well-defined 2 D space with both sources and sensors, and logical arguments suggest that it captures the essence of the full problem, however there are multiple obstacles to overcome in practice. One is the potentially very high dimensionality of the source space. As an example, suppose that sources are constrained to, say, 15000 positions on the cortical surface, allowing them to have an arbitrary orientation with degree resolution would entail searching over a space of size greater than 10^{10}, further augmented if 3D positions, or multiple or extended sources are allowed. The impact of noise and various imprecisions in real data is another concern. Thus, success is uncertain.

As proof-of-concept, the method was applied to 274-channel MEG data recorded in an auditory oddball paradigm (the same data as used for the Brainstorm introductory tutorial, https://neuroimage.usc.edu/brainstorm/Tutorials, Tadel et al 2011). Figure 12 shows gain maps associated with sources responsible for the auditory evoked response (A and B, left and right hemispheres respectively), the button press (C), and background alpha activity (D).

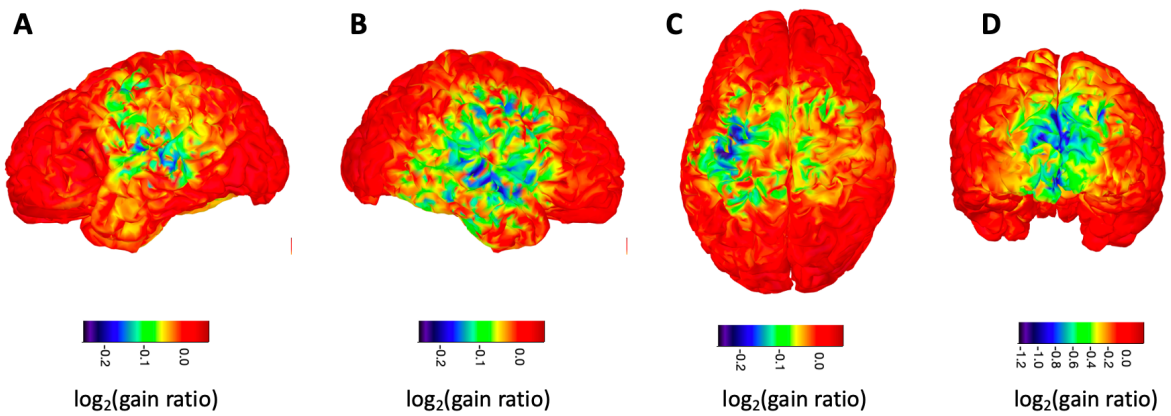

Fig. 12. Gain maps for null filters relative to an auditory evoked response (A and B), button press (C) and alpha activity (D). The color scale represents the base-two logarithm of the ratio between the absolute value of source-to-sensor gain of selected columns of the JD analysis transform matrix, normalized by the average value over all columns.

More precisely: to obtain panels A and B, MEG data in response to the auditory stimuli were epoched and processed by the JD algorithm to find a transform matrix that isolates, within its first few columns, stimulus-locked components of cortical activity (as in Sect. 4.5). Based on

a source/forward model appropriate for these MEG data calculated by Brainstorm (~45000 dipoles orthogonal to the cortical surface), source-to-sensor gain maps were calculated for each column of this transform matrix. The absolute value of the gain was averaged over all columns except the first three, and normalized by dividing by the absolute value of gain over all columns. This normalization is required to factor out the dependency of gain on depth. The base-two logarithm of this ratio, plotted as a color scale in Fig. 12 A and B, shows relatively small values in temporal regions bilaterally, consistent with the known position of auditory fields. Panel C was obtained in a similar fashion from data in response to button presses, showing a left-lateralized location consistent with motor activity. Panel D was obtained based on the JD algorithm biased to isolate components dominated by power near 10 Hz (de Cheveigné and Arzounian 2015). Ten components were assumed alpha-dominated, the plot is based on the remaining columns (11 and beyond). The location, clearly occipital, is consistent with alpha activity.

The consistency of these patterns with expectations based on the literature suggests that they capture something about the topography of actual cortical sources. However, utmost caution is required. A major concern is the mismatch between the model source space (single dipoles orthogonal to the cortical surface) and the real source space (with characteristics unknown). From earlier simulations, we might have expected a null (zero gain) at a particular position indicating the source location. Instead, we see a spatially-extended pattern of mildly reduced gains (note the scales on the color bars). The fact that gain is nowhere zero implies that *the source is not within the model source space*, in other words (as one might have guessed) the cortical source is not a single dipole normal to cortical surface. Instead, the patterns in Fig. 12 might be understood as reflecting a measure of similarity between the actual source and model dipole sources as a function of their position in space, viewed through the prism of the forward model and null filter.

In any case, it is important to realize that such patterns *do not represent the spatial pattern of brain activity*. While it is tempting to see the spatially-extended patterns of Fig. 12 as indicative of a spatially-extended source, that interpretation would be very unwise at this stage. Much more work is required, for example using simulations, before such interpretations can be made safe. This is food for future forays, but Fig. 12 suggests is that those efforts might be fruitful.

5.4.2 Relations with other methods

Another future direction is a detailed comparison with the many other methods reported in the literature, at a theoretical or empirical level. Of particular interest is the possibility that known equivalences between methods (e.g. Mosher et al 2003; Baillet et al 2010; Hauk et al 2022) might be extended to include this approach. At this point, this paper's claim for originality is in exposition and interpretation, rather than algorithmic novelty or superiority, but some discrepancies deserve resolution, for example concerning concepts such as virtual electrode, point spread, resolution matrix, and so on (Liu et al 2002; Samuelsson et al 2021; Hauk et al 2022) that seem to have no equivalent here.

5.4.3 Algorithmic and implementation issues

Localization requires *search* within the parameter space of the source model. The simulations here used exhaustive search within the disc based on a 500 × 500 grid, but this approach may be unwieldy for more complex spaces, for example within 3 D space, with arbitrary dipole orientation, or with complex shapes (e.g. patches instead of points). More efficient search is likely possible, exploiting constraints of the source model and the continuous and differentiable properties of the forward model, but this remains to be developed.

Conclusions

The main contribution of this paper is towards a better understanding of issues raised by linear analysis of multichannel brain data, based on a factorization into algebraic, geometrical, physical, computational, and representational factors, and with the help of a – hopefully – intuitive set of simulations based on a toy model that captures the essence of the problem. A first conclusion is that the appealing metaphor of a “virtual electrode” is best retired, as it is far from what can be achieved and easily misunderstood with possibly negative consequences. A second conclusion is that it is fruitful to reason in terms of *cancellation* of specific sources rather than their enhancement. Based on that, a third conclusion is that it is possible to *localize* brain sources very accurately within the parameter space of the source model, which may map directly to physical space for simple models such as a single dipolar source. Accurate localization and visualization (imaging) depend on a sequence of model-constrained computations that span from the signal and anatomy to the final image. Cancellation is central to this approach: cancellation of competing sources allows for optimal observation of the time course of a source of interest, while cancellation of a source of interest allows accurate

localization of that source. Based on theoretical arguments, simulations, and preliminary results (e.g. Fig. 12) we might expect a wide range of applications, scientific and clinical. If so, this paper only scratched their surface.

Acknowledgments

Malcolm Slaney and Lucas Parra gave useful comments on earlier drafts.
This work was supported by grants ANR-10-LABX-0087 IEC, ANR-10-IDEX-0001-02 PSL, and ANR-17-EURE-0017

References

Alberto, G. E., Stapleton-Kotloski, J. R., Klorig, D. C., Rogers, E. R., Constantinidis, C., Daunais, J. B., & Godwin, D. W. (2021). MEG source imaging detects optogenetically-induced activity in cortical and subcortical networks. *Nature Communications*, *12*(1), 5259. https://doi.org/10.1038/s41467-021-25481-y

Anemüller, J. (2007). Maximization of Component Disjointness: A Criterion for Blind Source Separation. In M. E. Davies, C. J. James, S. A. Abdallah, & M. D. Plumbley (Eds.), *Independent Component Analysis and Signal Separation* (Vol. 4666, pp. 325–332). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-74494-8_41

Baillet, S., Mosher, J.C., and Leahy, R.M. (2001) Electromagnetic brain mapping. IEEE Signal Processing Magazine, 18(6):14–30.

Baillet, S. (2010) The Dowser in the Fields: Searching for MEG Sources, in MEG, an Introduction to Methods, edited by Hansen, P.C., Kringelbach, M.L., Salmelin, R., Oxford University Press, 83-123.

Cao, M., Galvis, D., Vogrin, S. J., Woods, W. P., Vogrin, S., Wang, F., Woldman, W., Terry, J. R., Peterson, A., Plummer, C., & Cook, M. J. (2022). Virtual intracranial EEG signals reconstructed from MEG with potential for epilepsy surgery. *Nature Communications*, *13*(1), 994. https://doi.org/10.1038/s41467-022-28640-x

Cardoso, J.F., 2001. The three easy routes to independent component analysis; contrasts and geometry. Proc. Int. Conf. on Independent Component Analysis and Blind, Source Separation (ICA01), pp. 1–6.

Coelli, S., Medina Villalon, S., Bonini, F., Velmurugan, J., López-Madrona, V. J., Carron, R., Bartolomei, F., Badier, J.-M., & Bénar, C.-G. (2023). Comparison of beamformer and ICA for dynamic connectivity analysis: A simultaneous MEG-SEEG study. *NeuroImage*, *265*, 119806. https://doi.org/10.1016/j.neuroimage.2022.119806

de Cheveigné, A., Parra, L. (2014) Joint decorrelation: a versatile tool for multichannel data analysis. *NeuroImage*, 98, 487-505, doi: 10.1016/j.neuroimage.2014.05.068.

de Cheveigné A, Arzounian D (2015) Scanning for oscillations, *Journal of Neural Engineering*, 12, 066020, DOI: http://iopscience.iop.org/article/10.1088/1741-2560/12/6/066020/meta.

Engels, M. M. A., Hillebrand, A., van der Flier, W. M., Stam, C. J., Scheltens, P., & van Straaten, E. C. W. (2016). Slowing of Hippocampal Activity Correlates with Cognitive

Decline in Early Onset Alzheimer's Disease. An MEG Study with Virtual Electrodes. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00238

Fuglsang, S. A., Dau, T., & Hjortkjær, J. (2017). Noise-robust cortical tracking of attended speech in real-world acoustic scenes. *NeuroImage*, *156*, 435–444. https://doi.org/10.1016/j.neuroimage.2017.04.026

Hauk, O., Stenroos, M., & Treder, M. S. (2022). Towards an objective evaluation of EEG/MEG source estimation methods – The linear approach. *NeuroImage*, *255*, 119177. https://doi.org/10.1016/j.neuroimage.2022.119177

Helmholtz H. (1867). Handbuch der Physiologischen Optik (English tranl.: 1924 JPC Southall as Treatise on Physiological Optics) Voss.

Juárez-Martinez, E. L., Nissen, I. A., Idema, S., Velis, D. N., Hillebrand, A., Stam, C. J., & van Straaten, E. C. W. (2018). Virtual localization of the seizure onset zone: Using non-invasive MEG virtual electrodes at stereo-EEG electrode locations in refractory epilepsy patients. *NeuroImage: Clinical*, *19*, 758–766. https://doi.org/10.1016/j.nicl.2018.06.001

Litvak, V., Jha, A., Eusebio, A., Oostenveld, R., Foltynie, T., Limousin, P., Zrinzo, L., Hariz, M. I., Friston, K., & Brown, P. (2011). Resting oscillatory cortico-subthalamic connectivity in patients with Parkinson's disease. *Brain*, *134*(2), 359–374. https://doi.org/10.1093/brain/awq332

Litvak, V., Mattout, J., Kiebel, S., Phillips, C., Henson, R., Kilner, J., Barnes, G., Oostenveld, R., Daunizeau, J., Flandin, G., Penny, W., & Friston, K. (2011). EEG and MEG Data Analysis in SPM8. *Computational Intelligence and Neuroscience*, *2011*, 1–32. https://doi.org/10.1155/2011/852961

Liu, A. K., Dale, A. M., & Belliveau, J. W. (2002). Monte Carlo simulation studies of EEG and MEG localization accuracy. *Human Brain Mapping*, *16*(1), 47–62. https://doi.org/10.1002/hbm.10024

Lütkenhöner, B., 1998. Dipole source localization by means of maximum likelihood estimation I. Theory and simulations. Electroencephalogr. Clin. Neurophysiol. 106 (4), 314–321.

Millman, R. E., Prendergast, G., Kitterick, P. T., Woods, W. P., & Green, G. G. R. (2010). Spatiotemporal reconstruction of the auditory steady-state response to frequency modulation using magnetoencephalography. *NeuroImage*, *49*(1), 745–758. https://doi.org/10.1016/j.neuroimage.2009.08.029

Samuelsson, J. G., Peled, N., Mamashli, F., Ahveninen, J., & Hämäläinen, M. S. (2021). Spatial fidelity of MEG/EEG source estimates: A general evaluation approach. *NeuroImage*, *224*, 117430. https://doi.org/10.1016/j.neuroimage.2020.117430

Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., & Leahy, R. M. (2011). Brainstorm: A User-Friendly Application for MEG/EEG Analysis. *Computational Intelligence and Neuroscience*, *2011*, 1–13. https://doi.org/10.1155/2011/879716

Van Veen, B. D., Van Drongelen, W., Yuchtman, M., & Suzuki, A. (1997). Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Transactions on Biomedical Engineering*, *44*(9), 867–880. https://doi.org/10.1109/10.623056