version 14 April 2025

Virtual electrode or virtual scalpel?

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 the concept critically, and suggests 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, a spatial filter can be found that suppresses up to $J - 1$ sources while still retaining some information about other sources. This property can be used to improve the signal to noise ratio of observations, and also to accurately localize them as explored in a companion paper.

1. Introduction

The term "virtual electrode," applied to non-invasive multi-channel recordings of brain activity, refers to a spatial filter designed to enhance a particular source position. Given a recording from an array of 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 "virtual electrode" is justified by the fact that, like a real electrode, the filter targets a particular *spatial position*, and produces a single *time series*. By adjusting the filter, a virtual electrode is positioned – virtually – anywhere within the brain, with a spatial pattern of gain that mimics – we hope –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. Juárez-Martinez et al 2018; Alberto et al 2021; Cao et al 2022, Velmurugan et al 2022; Coelli et al 2023).

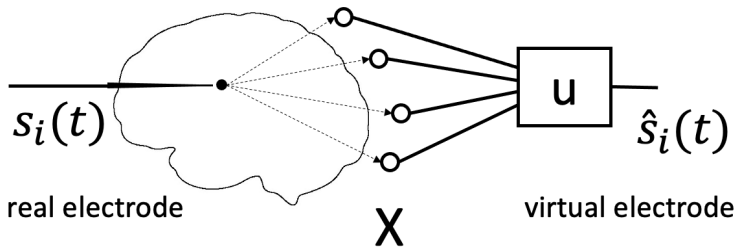

Fig. 1. A real electrode (left) is selectively sensitive to brain sources near its tip, and blind to distant sources. A virtual electrode (right) attempts to reproduce this property by applying a spatial filter to the multichannel observation matrix $\mathbf{X} = \{\boldsymbol{x}_j(\boldsymbol{t})\}$.

The prospect of clinical applications is appealing, and constitutes a drive for efforts to refine the methods and extend their applicability. A hospital administrator that hesitates at the cost of installing and running a MEG (magnetoencephalography) system might be swayed by the perspective saving 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 indeed replace a real electrode.

A simple mathematical argument tells us that real and virtual electrodes are not the same. A real electrode can be placed anywhere in the brain, potentially picking up activity different from any other location. If a real electrode could be placed in every one 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 with a rank that could potentially reach I. In contrast, with J sensors or electrodes, a set of I virtual electrodes can result in a data matrix of rank no greater than J (e.g. 64 or 128 for a typical EEG system). Since $J \ll I$, these data matrices are necessarily different, and thus 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". Conversely, 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: they reflect the fundamental non-uniqueness of the inverse problem (Baillet 2010). As in other scientific fields – and indeed perception itself (Helmholtz 1867) – additional information is required in the form of an *internal model* constrained by the observations. This approach works well, witness the success of EM imaging methods in neuroscience. Nevertheless, we may (indeed we should) feel uneasy using the outcome of that process as *evidence* to ground a further inference, because it is unclear to what degree it reflects the observation or the model (Tarantola 2006). Is the sharp boundary of activity, apparent in the image, a reflection of the data, or of the prior? Conversely, if the image is blurry, is it due to a *method-related or equipment-related limitation* (that is amenable to progress) or a *fundamental limitation* (that is not). A goal of this paper is to help answer such questions or, at least, formulate them.

In sum, the metaphor of a virtual electrode is useful to convey the power of source localization and imaging approaches, and foster enthusiasm for them, but it carries with it the seeds of misunderstanding their limits, and whether those limits can be overcome by better skill, algorithms or hardware. This paper proposes a different metaphor, less enticing but possibly more fruitful, that of a “virtual scalpel”. While it is not possible, in general, to focus on a single source and extract it from within the recorded data, it is always possible to *silence* that source, and by doing so, in some situations, locate it with precision.

The aim of this paper is not to propose new methodogy, but rather to offer insight into a class of powerful methods that not allways fully understood because of the non-intuitive properties of high-dimensional signal spaces, and the hard-to-visualize nature of source spaces within the brain.

2. Concepts and methods

I first describe the theoretical background and 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 of a dipole). Finally, I attempt to draw conclusions applicable to sources within a real brain as measured for example using EEG or MEG.

2.1. 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 recorded data are assumed to reflect the activity of I sources within the brain via a linear mixing process:

$$\mathbf{X} = \mathbf{SM} \qquad (1)$$

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, so $I \gg J$. Noise sources are assumed to belong also to $\mathbf{S}$, hence no noise term in Eq. 1.

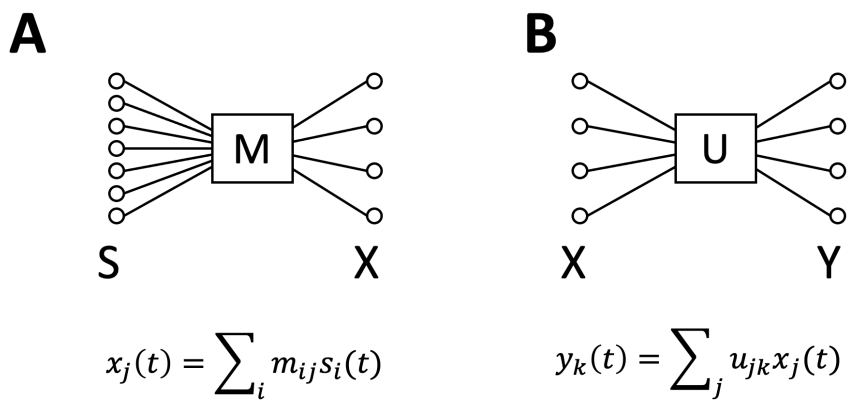

Fig. 2. A: Data model. The signal of each of the J observed channels is the weighted sum of I brain (or noise) source signals. B: Analysis model. The signal of each of the K components is the weighted sum of J observed channels.

2.2. Analysis model

We are usually interested in the activity of certain brain sources, i.e. one or more columns of **S**. 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 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**.

Ideally, we would like the analysis matrix **U** to be an "unmixing matrix" so that **MU=I**, the identity matrix. Each column of **Y** would then map to a source. Unfortunately, brain sources number in billions, whereas sensor channels range from tens to thousands, so **M** is not invertible and therefore recovering **S** is impossible. We must settle for the less ambitious objective of finding a transform that produces components that are somehow informative. For example, a column of **Y** might approximate the activity of a particular source, or the compound activity of several sources with a functional meaning such as an event-related potential (ERP).

The matrix **U** might be predefined, for example re-referencing, Laplacian, or current source density (CSD) transforms. Or it might be produced by a data-driven method such as principal component analysis (PCA), independent component analysis (ICA), or other blind source separation (BSS) methods. Each column of **U** constitutes a *spatial filter*, a vector of size J which applied to the observations, produces a single, univariate time series.

2.3. Vector spaces

It is helpful to see time series of neural activity, or recorded data, as belonging to a *vector space*, because concepts such as *subspace*, *span*, *dimensionality*, *linear dependence* and *orthogonality* allow for useful insight and convenient short-cuts. The brain source signals $s_i(t)$ span a signal space of dimension I, at most, although it might be smaller if some of the $s_i(t)$ are linearly dependent. The observed data $x_j(t)$ span a subspace of that space, of dimension of at most J. The mixing matrix **M** projects from the source space onto this subspace. Since $I \gg J$, a given observation x_j can reflect many different patterns of brain activity (Fig. 3). This is another manifestation of the non-uniqueness mentioned earlier.

Components (columns of **Y**) belong to the subspace spanned by **X**. Some analysis methods (e.g. PCA) result in columns of **Y** that are mutually uncorrelated, in which 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.

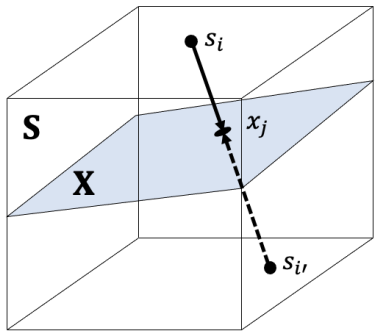

Fig. 3. The space spanned by the columns of the observation matrix **X** is a subspace of the space spanned by the brain signals **S**. Brain source signals are projected onto this subspace, and thus multiple sources may map onto the same observation.

2.4. Spatial filter, virtual electrode, null filter, zero set

A *spatial filter* is a vector of J weights $\{u_j\}$ to be applied to the observation channels. The gain between a source i and the filter output is:

$$g_i = \sum_j m_{ij}\, u_j, \qquad (2)$$

i.e. a weighted sum of the gains from that source to each sensor. One interpretation is that the filter "interpolates" between sensors or electrodes to create a new channel, hence the moniker "virtual electrode". Spatial filters too form a vector space, because any linear combination of filters is also a filter. Its dimension is J because there are J coefficients.

A *null filter* $\mathbf{u}$ for a source $\mathbf{s}_i$ is a filter that suppresses that source, i.e. the source-to-filter output gain is zero:

$$\sum_j m_{ij}\, u_j = 0, \qquad (3)$$

Every source admits null filters (as long as $J > 1$), a trivial example being a filter with all coefficients zero except for two channels j and j' such that $u_j/u_{j'} = -\, m_{ij}/\, m_{ij'}$. Since a weighted sum of null filters for a source is a null filter for that source, null filters for a source form a vector space, subspace of the space of spatial filters. Given the constraint imposed by

Eq. 3, the subspace of null filters for a source is of dimension $J-1$. A source has many null filters, thus, there are many ways to create a "virtual electrode" blind to that source.

Given a second source i', there exists likewise a subspace of null filters of dimension $J-1$ for that source. The null filters for *both sources* belong to the intersection of the null filter subspaces for each, i.e. a subspace of dimension $J-2$. Extending this argument, for J observation channels one can suppress at least $J-1$ sources (see Fig 6 for examples).

A real electrode is sensitive to sources near its tip, and insensitive to distant sources in the brain (Fig. 1, left). To reproduce that property, a spatial filter (Fig. 1 right) must put a null on each of those distant sources. 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, in general. This is yet another manifestation of the non-uniqueness mentioned earlier.

The *zero set* of a filter is the set of sources for which the gain is zero. The zero set of a null filter for a source includes that source (by definition), but based on the reasoning that led to Fig. 3, we expect there to be other positions for which the output is zero. In other words, the zero set may be extended in source space (see Figs. 4D and following for illustrations).

2.5. Data-driven analysis

So far, we did not say how analysis matrices $\mathbf{U}$ or spatial filters $\mathbf{u}$ are chosen. In some cases, they are predetermined, for example interpolation between electrodes, or Laplacian, in others they are calculated based on a data-driven method such as PCA, ICA, beamforming, or others.

Among data-driven methods, PCA stands out for three useful properties. First, the columns of the component matrix $\mathbf{Y}=\mathbf{XU}$ 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. A consequence of this third property is that the *last* principal component includes the *least* possible amount of variance. Indeed, its variance is zero if the data are rank-deficient. If at most $J-1$ sources were active, the last column of the PCA matrix is a *null filter* for them all.

Other methods offer different properties. ICA attempts to ensure that components are *statistically independent* according to some empirical measure (Cardoso 2001), in addition to

uncorrelated. Canonical correlation analysis (CCA) seeks linear combinations of columns of $\mathbf{X}$ that are *maximally correlated* with some other dataset $\mathbf{X}'$. Joint decorrelation (JD, also known as denoising source separation, DSS, or common spatial patterns, CSP), seeks to maximize the *ratio of variance* between the data and a filtered version of the data (see de Cheveigné and Parra 2014 and references therein). These methods share with PCA the property that columns of $\mathbf{Y}$ are mutually uncorrelated.

2.6. Existence vs identifiability

Data-driven analysis is a powerful tool to find an analysis matrix (or spatial filter) with desired properties, but it can fail. Failure can be of two sorts: (a) *the desired solution exists but we fail to find it*, for example due to overfitting, or geometrical inaccuracy, or lack of appropriate criteria, or (b) *the desired solution does not exist*, as a result of a fundamental constraint. In the former case, a new technique might overcome the limitation; in the latter case, it cannot.

2.7. Localization and imaging

The discussion so far made no reference whatsoever 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 in the brain, or within the parameter space of a model of the brain. However, by associating observations with a *source model* (that defines the range of expected source types and locations) and a *forward model* (that defines the gain between each source type/location and the sensors), it is possible to make inferences about the source geometry, within the limits imposed by non-uniqueness of the inverse problem (Baillet 2001, 2010). Popular methods 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). In a companion paper I suggest an additional approach dubbed *zero set imaging* (de Cheveigné 2024), however I do not dwell on localization or imaging here.

3. Results

3.1. A toy world

The previous concepts are somewhat abstract, and it may be hard to grasp the relations between *signal spaces*, the *parameter space* of a source model, and *physical space* (e.g. position and orientation of a dipole). 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 easy-to-visualize 2-D space. This section serves as an illustrated recapitulation of the previous section.

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 isotropically with distance as $1/d^2$ (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, its gain pattern (color gradient in Fig. 4B) would be selective for that particular source. Unfortunately, sensors constrained to the periphery are less helpful for selecting sources within the disc.

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* as was argued in Sect. 2.4. For example, the weighted sum of two sensors can be chosen to form a spatial filter with a position-dependent gain as illustrated in Fig. 4C, so as to obtain *a null filter* for a source at the position of the green cross. The gain is also zero for other positions (the *zero set*) visible as a black line, as was hinted to in Sect. 2.4.

Fig. 4C illustrates several of the concepts introduced in Sect. 2.4. First, for any source, there exists a null filter. Second, that filter suppresses sources at other locations (the zero set). Third, the gain is non-zero for sources situated in most other locations, i.e. the filter output is sensitive to most sources. Figure 5 illustrates a fourth point: there exist multiple null filters for a source.

Fig. 5 also serves to illustrate a point made in Sect. 2.5: null filters may be obtained by data-driven analysis. Here, a 1000×1 "source activity" time series was multiplied by a 1×5 source-to-sensor gain matrix corresponding to a source at the green cross, to produce a 1000×5 "data matrix" $\mathbf{X}$. This matrix, of rank 1, was submitted to PCA to obtain an analysis matrix, the first column of which captures all the variance of the data. Columns 1-4 are thus null filters for this source, as evident in the fact that their zero sets (black) all include that position. These four filters form a basis for the space of null filters: all null filters are linear combinations of these. As pointed out in Sect. 2.4, there are multiple ways to silence a source using spatial filtering.

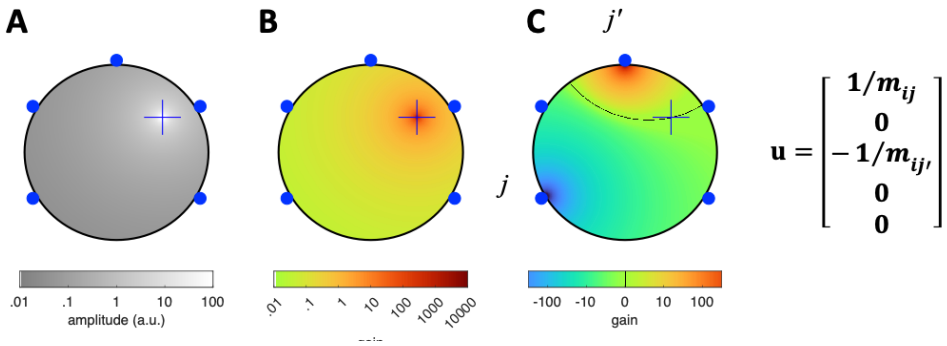

Fig. 4. A: Simple setup consisting of an array of sensors (blue dots) placed at the periphery of a disc. A source within the disc (cross) radiates within the plane with an amplitude that decreases with distance as $1/r^2$ (gray scale). B: Gain pattern for an electrode placed at the cross (color scale). C: Gain pattern for a spatial filter. This filter was designed to have zero gain at the cross ("null filter"). The black line indicates other locations of zero gain (the "zero set" for this filter).

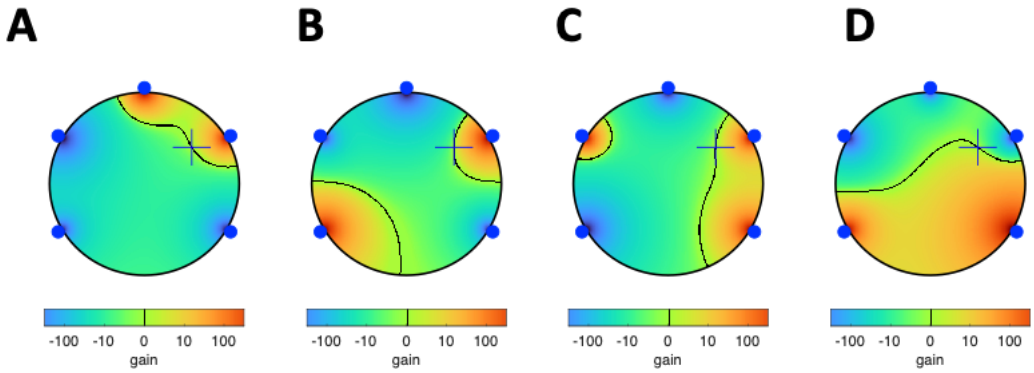

Fig. 5. Gain maps for spatial filters derived from columns 2 to 5 of the transform matrix of PCA applied to data from 5 sensors in response to a single source located at the green cross. These four filters are null filters for that source: the source (cross) is located within the zero set of each filter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

3.2. 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. For two sources the subspace is 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 (Fig. 6 C), *unique to a multiplicative constant*. Figure 6 D shows a similar example with 15 sensors and 14 randomly-placed sources; this filter too is unique.

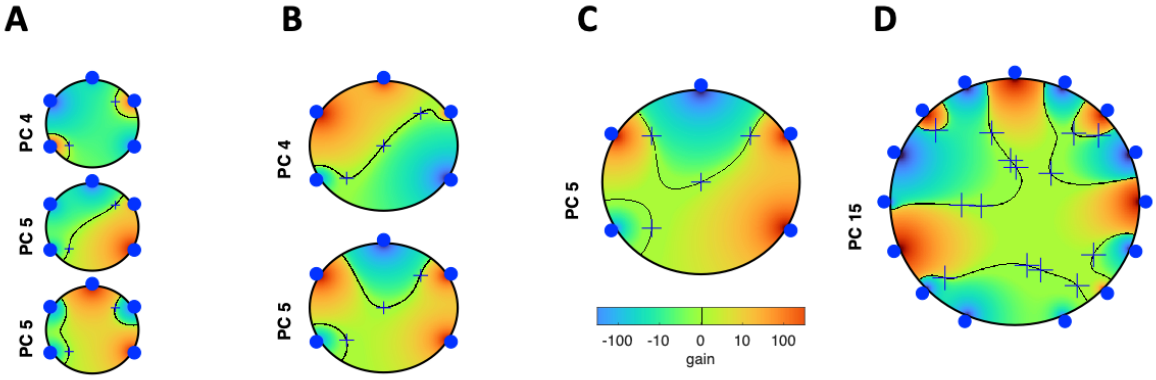

Fig. 6. Gain maps of null filters for 2 sources (A), 3 sources (B) and 4 sources (C), assuming 5 sensors. D: same for 14 randomly-placed sources and 15 sensors.

This illustrates an important conclusion: with J sensors one can silence up to $J-1$ arbitrarily-positioned sources. One cannot silence more, unless they all happen to fall within the zero set of a filter. The filters that silence K sources form a subspace of dimension $J-K$; if $K=J-1$, that subspace is of dimension 1, i.e. the filter is unique to a multiplicative factor.

3.3. A virtual electrode?

A virtual electrode should reproduce two properties of a real electrode: (1) it can boost the SNR (signal to noise ratio) of a target source relative to competing sources (by placing the "tip" near that source), and (2) it can be used to localize the source of a given signal by moving the tip around until that signal is found. I'll consider each in turn.

SNR enhancement. We can define "SNR enhancement ratio" as the ratio between the SNR of the filter output and the largest SNR over sensors. From Sect. 3.2, we know that there exists a filter for which this ratio is infinite, as long as the number of competing sources is $J-1$ or less, and unless the target source happens to be located within the zero set of that filter (black lines in Fig. 6C,D). If the number of competing sources is larger, we can only hope find a filter that maximizes this ratio.

How to find that solution? If we know the *location* i of the target in source space, and we have a source-to-sensor *forward model*, then we can apply the linear constrained minimum variance (LCMV) beamformer to find the desired spatial filter:

$$\mathbf{u} = (\mathbf{f}_i^{\top}\mathbf{C}^{-1}\mathbf{f}_i)^{-1}\mathbf{f}_i\mathbf{C}^{-1} \qquad (4)$$

where $\mathbf{C}$ is the covariance matrix, and $\mathbf{f}_i$ the source-to-sensor gain vector for that location (Van Veen et al 1997; Jaiswal et al 2020). The LCMV beamformer produces a filter $\mathbf{u}$ that cancels (or minimizes) variance from sources other than at that location.

This is illustrated in Fig. 7A for a toy world configuration with 5 sources: one target (circle) and four competitors (crosses). The target was simulated as ten cycles of a sinusoid (total duration 1000 samples), and the competing source signals as a 1000×4 matrix of uncorrelated Gaussian noise. The forward model was applied to transform source signals into a 1000×5 matrix of sensor signals. The covariance matrix of these data was calculated, and Eq. 4 applied with $\mathbf{f}_i$ appropriate for the target position. The gain is indeed zero at the locations of the competitors (crosses), and thus the SNR enhancement ratio is infinite. Not unexpectedly, given the previous discussion, the gain pattern of Fig. 7A is the same as that in Fig. 6C, to a factor.

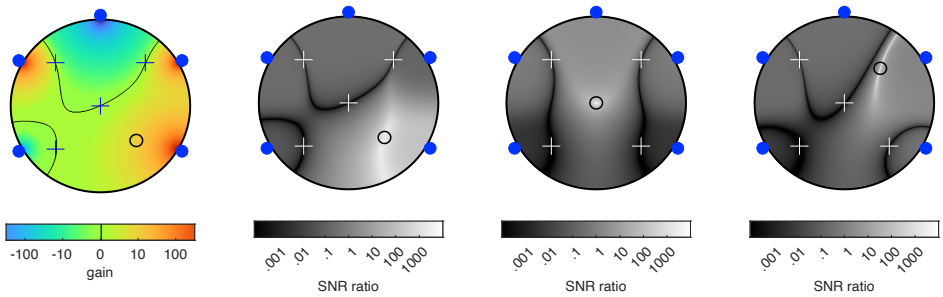

Fig. 7. A: Gain map for the filter produced by a beamformer applied to sensor signals produced from one target source (circle) and four competitors (crosses), when the beamformer is aimed at the target (circle). B-D: SNR of the target signal at the output of the filter produced by a beamformer as a function of the location at which the beamformer is aimed, for three target positions.

If we do *not* know the target location, but we can distinguish the target signal (here a sinusoid with period 100 samples) from that of the competitors (here Gaussian noise), we can scan the source parameter space (here the disc), applying the beamformer to each location, and select the location that yields the largest SNR. This is illustrated in Fig. 7B: the SNR ratio is large (white) for locations within a swath of the disc near the target (circle). Panels C and D show similar simulations for other positions of the target.

As an alternative to scanning with a beamformer, we can use a data-driven algorithm such as JD (de Cheveigné and Parra 2014) to directly find a filter that maximizes the SNR. This was simulated using the same 1000×5 matrix of sensor signals. JD was applied, using a criterion of maximum variance of the trial-averaged signals, and the first column of the analysis matrix

used as a filter. The gain map, plotted in Fig. 8A, is identical to that in Fig. 7A (to a factor), itself identical to that in Fig. 6C (to a factor), illustrating once again the unicity of this filter. The appeal of JD over beamforming is that it does not involve a forward model and does not require scanning.

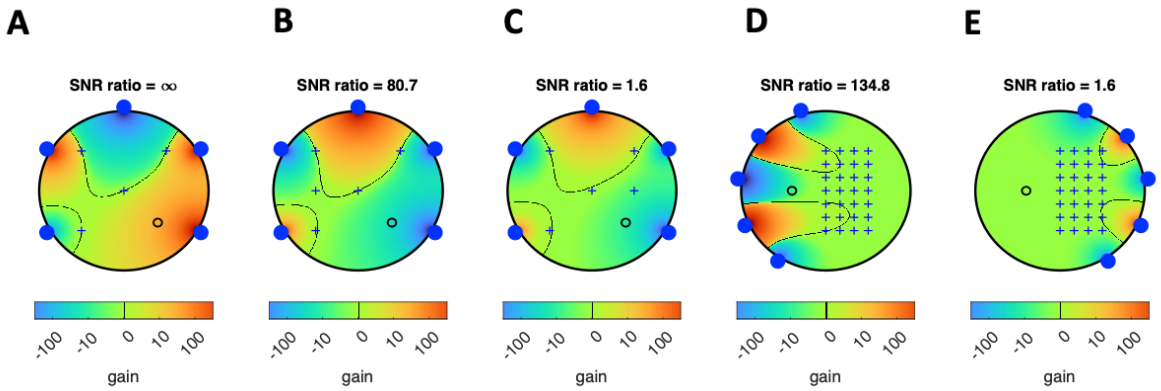

Fig. 8. Gain maps spatial filters ("virtual electrodes") created using the JD algorithm (see text) for several configurations of sensors and sources (circle: target, crosses: competitors).

The same methods can be applied with more than $J-1$ competing sources, but in general the SNR enhancement ratio is smaller. The ratio depends strongly on the configuration (Figs. 7B-E), ranging from small (panels C or E) to larger (panels B or D). For a given number of competing sources, it can differ depending on their location (compare panels B and C), and for the same source locations and number of sensors, it can depend on the configuration of the sensors (compare panels D and E). The SNR-enhancement power of such "virtual electrodes" is not easy to predict. Overall, their gain patterns bear little resemblance with that of a real electrode (Fig. 4B).

Source localization. The second useful property of a real electrode that we might like to reproduce is the ability to localize the source of a particular signal by moving around its tip. This property can be emulated with the help of a beamformer as illustrated in Fig. 7B-D. The SNR (measurable in terms of periodic-to-noise variance ratio) peaks in a region near the target, however in some cases (e.g. Figs. 7B,D) there is quite a wide spread. Beamforming is one of several popular approaches to source localization (Baillet 2001, 2010).

3.4. A virtual scalpel?

Whereas the goal of assigning a high gain to a target and zero gain to all competitors cannot be attained (except for very particular configurations), that of assigning *zero gain to a target* and

non-zero gain for most other sources can be achieved perfectly (Fig. 4C and Figs. 5-8). This suggests a strategy by which sources are successively removed from the data by applying successive stages of linear filtering. This idea can be embodied in several ways.

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

$$\mathbf{D} = \mathbf{U}\mathbf{E}\mathbf{U}^{-1} \qquad (5)$$

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:

$$\widehat{\mathbf{X}} = \mathbf{X}\mathbf{D}. \qquad (6)$$

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, if further analyses call for a forward model, it needs to be adjusted for effects of the denoising matrix (but see Hipp and Siegel 2015). Examples of this approach are the common use of ICA to isolate and remove ocular artifacts, or JD to remove power line artifacts.

A second approach is applicable in the case where the response of interest is made up of multiple spatio-temporal components reflecting the existence of multiple sources with different spatial signatures and different time-courses. This situation is common in the analysis of stimulus-evoked responses, for which latencies and/or time constants can differ between hemispheres and/or primary and higher-order fields. This is illustrated in Fig. 9 for a 5-source toy-world configuration. Two target sources (with locations indicated as circles and distinct but correlated time courses) were mixed with three competing sources (with locations indicated as crosses and time courses drawn from independent Gaussian processes). JD analysis resulted in two components, each of which appears to be a mixture of the two targets (Fig. 9B), as the corresponding filters both have non-zero gain at both source locations (Fig. 9C,D). However, the fact that the target source waveforms are staggered can be used to perform a second stage of JD analysis that allows on of the sources to be cancelled, thus isolating the other (Figs. 9E-H). For panels E,F, the JD analysis was designed to maximize power in the interval 1-100 samples relative to the rest, for panels G,H it was designed to maximize power in the interval 201-400 samples, i.e. intervals where one or the other source is quiescent.

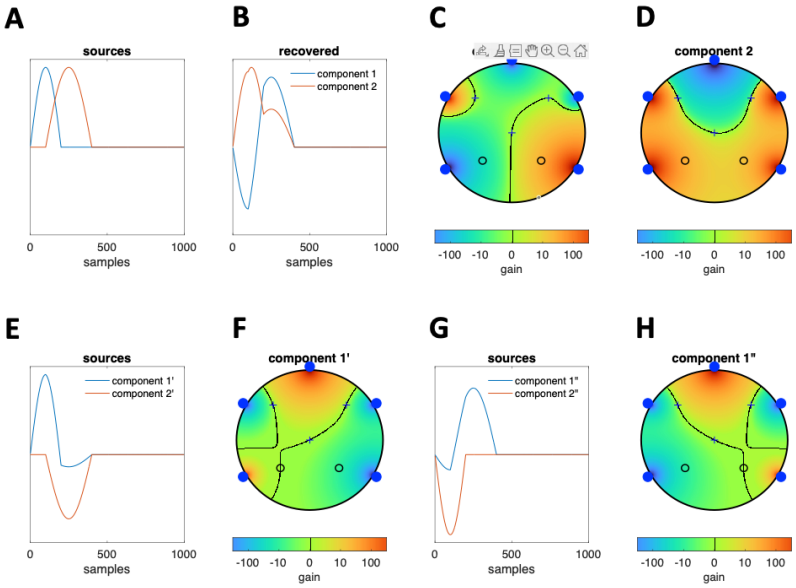

Fig. 9. A: Time course target sources (with locations indicate as circles in panels C,D). B: Time courses of components retrieved by JD. C,D: Gain maps of filters associated with the first two components in B. E: Time courses of components rotated by JD to isolate the later source. F: Gain map of corresponding filter. G: Time courses of components rotated by JD to isolate the earlier source. H: Gain map of corresponding filter.

A third approach exploits the zero sets of null filters for a source to localize that source. This is described in a companion paper (de Cheveigné 2024) and will not be discussed further here. In each of these approaches, the spatial filter can be likened to a "virtual scalpel" with which individual sources are cancelled for the purpose of *characterizing* them, and *suppressing* them to reveal other sources.

In sum, these "toy world" simulations illustrated properties of spatial filters relevant for understanding and evaluating the concept of "virtual electrode". An important observation is that it is in general *not* possible to reproduce the spatially selective gain pattern of a real electrode (Fig. 4B), although under some conditions the filter may mimic its SNR enhancement properties (e.g. Fig. 8), and its source localization properties (Fig. 7B-D). A spatial filter cannot, in general, isolate a single source among many, but it can perfectly *suppress* one source among many. This property may contribute to parse the compound multivariate response into more meaningful "sources" and determine their location (de Cheveigné 2024).

3.5. Analysis of real MEG data

The "toy world" served to elucidate some basic properties, but one can wonder how the conclusions translate to real brain recordings. The concepts discussed in Sect. 2 are all applicable, but their effects are harder to visualize because of the complexity of the signal space (here 274 channels) and the source space (typically 3D position plus orientation anatomically constrained to a convoluted surface).

Publicly-available data from the BrainStorm introductory tutorial consisted of 274-channel MEG signals recorded in response to repeated auditory stimuli (tones), together with spatial and anatomical information (MRI scan co-registered with sensor positions) for the subject. The BrainStorm software suite was used to calculate a forward model sampled at 405006 dipole source locations on the cortical surface, with orientations normal to the surface.

MEG data sampled at 600 Hz were high-pass filtered (0.5 Hz cutoff) and smoothed (10-sample boxcar) and their dimensionality reduced to 40 components with the SCA (shared component analysis, similar to PCA) algorithm (de Cheveigné 2021). In a first step, a data-driven algorithm (JD, de Cheveigné and Parra 2014) was used design a set of filters that maximize the ratio of variance of trial-averaged to raw data which is relatively large for the first two components (Fig. 10A). The subsequent components are ignored here; their smaller scores could reflect poorer SNR or brain processes that are less perfectly reproducible from one trial to the next. The choice of cutoff (2) is arbitrary.

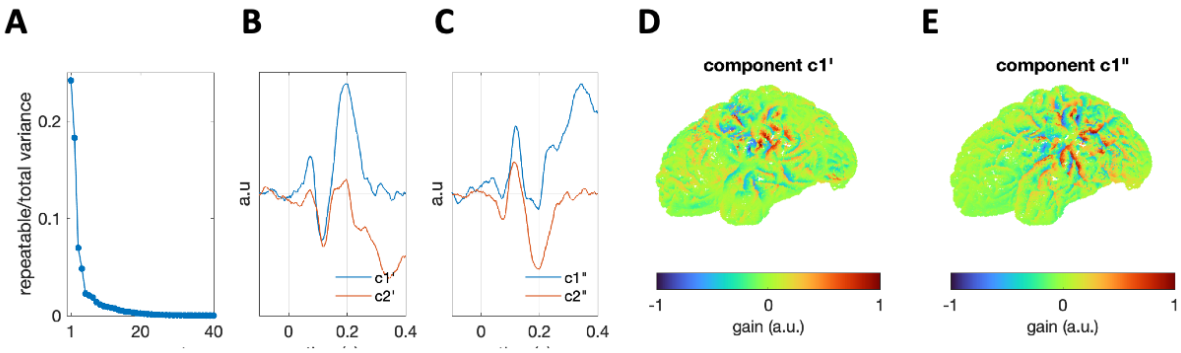

Fig. 10. MEG response to repeated tone stimulation. A: Ratio of repeated to total variance for each component of a JD analysis. B: Trial-averaged time course of the first two components, rotated so that the first (blue) isolates the early response (<50 ms). C: Trial-averaged time course of the first two components, rotated so that the first component isolates the late response (>300 ms). D: Gain map of the filter associated with the early component. E: Gain map of the filter associated with the late component.

In a second step, the first two components were submitted to a JD analysis designed to maximize the early response (between 50 and 90 ms) relative to late (90 to 400 ms). The time-averaged components from this second analysis are plotted in Fig. 10B (signs and amplitudes are arbitrary). The first component (blue) includes an early deflection that is absent from the second (in addition to a series of later deflections). A similar analysis was performed to maximize the late response (300 to 400 ms) relative to early (0 to 300 ms). The time-averaged components from this second analysis are plotted in Fig. 10C. The first component (blue) includes a late deflection that is absent from the second (in addition to a series of earlier deflections). These two analyses, analogous to those of Fig. 9E-H, are relevant *under the assumption* that there exist two main spatio-temporally distinct sources, one with short latency (<90 ms) and another with long-lasting activity (>300 ms). That this is, at best, an approximation is suggested by the possible presence of addition components (e.g. 3 and 4 of the initial JD analysis) and the multiple intermediate deflections (between 90 and 300 ms) that seem to partially overlap between these "sources". Whether this approximation is useful depends on how it fits with other empirical observations, and whether it contributes to their understanding. If it is not, the best we can say is that these components *span a subspace* that contains stimulus-evoked components.

The red lines in panel C resemble the blue in B and vice-versa, suggesting that the time courses of these "early" and "late" sources are roughly decorrelated, which is not to be expected in general; indeed, it was not the case for the simulation of Fig. 9. The gain patterns for the spatial filters that best select these "sources" are shown in Fig. 9D,E. Considering the first (panel D), it appears to show large values (positive or negative) centred over temporal regions, as expected for an auditory response, but also over regions quite distant. It is tempting to see this extended gain pattern as reflecting the distribution of cortical sources that gave rise to the response from which the filter was calculated. To show that this is not the case, the time series of the stimulus-locked response was regressed out of the MEG data, and used to simulate a single dipolar source (located at the position of maximum gain of the previous filter), that was then added to the background MEG. The JD algorithm was applied again to obtain a filter that optimally extracts the activity of this source (now a single dipole). The gain pattern of the resulting filter (not shown) is indistinguishable from that plotted in Fig. 10A, suggesting that the temporal extent of the gain pattern does *not* reflect the distributed nature of cortical sources.

Rather, the extended pattern can be interpreted as follows. First, note that the gain pattern of a filter is a weighted sum of gain patterns associated with each of the sensors. These live in a

space of dimension 274, and thus we cannot expect a distinct focal gain pattern for every one of the 405006 potential source positions. The non-focal nature of filter gain patterns is thus to be expected, as yet another manifestation of the non-unicity of the inverse problem. Second, consider that the filter produced by the JD algorithm attempts only to optimize the SNR of the repeatable target. For this, it may assign (a) a high gain to sensors with high SNR (e.g. close to the source), and (b) non-zero gain to additional sensors so as to *subtract* competing source activity (e.g. sensors close to competing sources). Furthermore, (c) it is free to assign arbitrary gain to regions that are relatively quiescent, as these will not affect the SNR. Thus, the distribution of gains is likely to not match that of target brain source activity.

The gain pattern of the filter that best selects the late source (Fig. 10E) is likewise spatially extended. From the analysis of Sect. 2 and the simulations of Sects. 3.1-3.4, we expect similar gain patterns whatever the analysis technique (e.g. JD or beamforming). We expect the gain pattern to be largely determined by the distribution of *competing* source activity, as SNR enhancement depends mainly on cancelling those sources. With beamforming, the distribution of competing sources is largely captured by the covariance matrix ($\mathbf{C}$ in Eq. 4); if that matrix is estimated on a segment for which a particular source is quiescent, the filter may not be effective in removing it over segments for which it is active (this consideration is germane to the decision to estimate it over a "baseline" segment distinct from the "response" segment). Neither gain pattern resembles the spatially-selective pattern characteristic of an "electrode" (as was schematized in Fig. 4B).

4. Discussion

Linear methods are the "work-horse" of EM brain data analysis. In a non-invasive method such as EEG or MEG, the observations result from quasi-instantaneous linear mixing. Thus, it makes sense to consider linear methods to "un-mix" them, with the caveat that the sparsity of the observations prevents perfect unmixing (Baillet 2001, 2010).

A well-chosen spatial filter can increase the SNR of sources of interest far beyond that available at any of the sensors, which motivates the term "virtual electrode" to describe such a filter. The term is attractive, but it carries two downsides. First, it is tempting to over-state its power, and consider non-invasive techniques as an alternative to invasive exploration (e.g. SEEG or ECoG for pre-surgical exploration in epilepsy). Second, the variety and complexity of methods preserve the hope that the filter can be further perfected, despite performance being limited by theoretical constraints. The purpose of this paper is to make these aspects easier to understand.

A first result is that a "virtual electrode" differs fundamentally from a real electrode. Many arguments point to this, including non-uniqueness of the inverse problem (Baillet 2001, 2010), dimensionality-based arguments developed in the Introduction, and the concrete toy world examples described in Sect. 3 (compare the gain pattern in Fig. 4B to those in Figs. 5-8), or the illustration with a real MEG source model (Fig. 9). A virtual electrode is much more "promiscuous" than a real electrode.

A second result is that, in a subset of situations, a benefit similar to an invasive electrode can be obtained in terms of SNR gain (e.g. Fig. 8). Roughly speaking, these situations correspond to those for which fewer than J competing sources are prominent.

A third result is that it is always possible to perfectly *cancel* a source. Indeed, with J sensors we can perfectly cancel up to $J - 1$ sources. In general, it is *not* possible to cancel more, unless they all fall within the zero set of a filter (black lines in Figs. 4-8). The filter that cancels $J - 1$ sources is unique (to a factor) and, more generally, filters that cancel $K < J$ form a subspace of dimension $J - K$. Herein lies the real power of linear filtering. A recording of brain activity can be successively stripped of the contribution of multiple sources *as long as the appropriate filters can be found*. Finding them might be challenging, but it is useful to know that they exist.

The ability to cancel sources can assist source localization in at least two ways. One is the well-known approach of scanning the source parameter space with a beam-former (as illustrated in Fig. 7B-D). Here, cancelling $J - 1$ competing sources assists the localization of the J-th target source. The other is the zero-set method described by de Cheveigné (2024), in which the target source location is determined as the intersection of zero sets of filters that cancel it.

The parameter J (number of sensor channels) determines the dimension of the signal space, and thus the number of competing sources that can be cancelled: the more, the better, with the caveat that the appropriate solution may be harder to find due to overfitting (more below). This is an argument in favor of high-density recording systems, the utility of which is sometimes questioned (e.g. Montoya-Martinez et al 2021), with the caveat that sensor-specific instrumental noise sources add to the number of competing sources.

For a given number of sensors, the effective dimensionality can be augmented in at least three ways. One is to split the data into different time intervals (allowing for a larger dimensionality ultimately bounded by $T \times J$). Another is to split the data over a filter-bank, or more generally a bank of convolutive kernels. For example, a set of F time shifts allow search within a space of dimension $F \times J$ for solutions that consist of multichannel FIR (finite impulse response)

filters. A third approach is to apply non-linearities as in the QCA (quadratic component analysis) method of de Cheveigné (2012). In each case, the effective dimensionality of the data is augmented, and with it the number of parameters of the solutions and thus the risk of overfitting.

"Overfitting" refers to the situation when (a) a solution exists and (b) it cannot be found because the statistics that it is based upon (e.g. means or covariance matrices) are unreliable due to paucity of data to constrain the available parameters. It is useful to distinguish this situation from that where no solution exists. An aim of this paper was to better delineate the boundaries within which solutions exist, as distinct from our ability to find them.

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 may be 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 zero 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 zero set, increasing the likelihood that at least one of them will preserve the sources of interest. This is another argument in favor of high-density recording devices.

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 inexact 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, as a method to retrieve activity of interest, and to localize it. Cancellation of competing sources allows for optimal observation of the time course of a source of interest, while cancellation of a source of interest man allow accurate localization of that source.

Acknowledgments

Malcolm Slaney, Lucas Parra and Israel Nelken 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

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. (2012) Quadratic component analysis, NeuroImage, 59, 3838-3844, https://10.1016/j.neuroimage.2011.10.084.

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 (2021). Shared Component Analysis. *NeuroImage*, 226, 117614, https://doi.org/10.1016/j.neuroimage.2020.117614.

de Cheveigné, A. (2024) What and Where in Electromagnetic Brain Imaging (2024), *bioRxiv*, https://www.biorxiv.org/content/10.1101/2024.04.17.589904.

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.

Hipp, J. F., & Siegel, M. (2015). Accounting for Linear Transformations of EEG and MEG Data in Source Analysis. *PLOS ONE*, *10*(4), e0121048. https://doi.org/10.1371/journal.pone.0121048

Jaiswal, A., Nenonen, J., Stenroos, M., Gramfort, A., Dalal, S. S., Westner, B. U., Litvak, V., Mosher, J. C., Schoffelen, J.-M., Witton, C., Oostenveld, R., & Parkkonen, L. (2020). Comparison of beamformer implementations for MEG source localization. *NeuroImage*, *216*, 116797. https://doi.org/10.1016/j.neuroimage.2020.116797

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., 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

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

Montoya-Martínez, J., Vanthornhout, J., Bertrand, A., & Francart, T. (2021). Effect of number and placement of EEG electrodes on measurement of neural tracking of speech. *PLOS ONE*, *16*(2), e0246769. https://doi.org/10.1371/journal.pone.0246769

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

Tarantola, A. (2006). Popper, Bayes and the inverse problem. *Nature Physics*, *2*(8), 492–494. https://doi.org/10.1038/nphys375

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

Velmurugan, J., Badier, J.-M., Pizzo, F., Medina Villalon, S., Papageorgakis, C., López-Madrona, V., Jegou, A., Carron, R., Bartolomei, F., & Bénar, C.-G. (2022). Virtual MEG sensors based on beamformer and independent component analysis can reconstruct epileptic activity as measured on simultaneous intracerebral recordings. *NeuroImage*, *264*, 119681. https://doi.org/10.1016/j.neuroimage.2022.119681