version 19a Sept 2023

Virtual electrode, or virtual scalpel?

Alain de Cheveigné
ENS / CNRS

1. 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, and with appropriate source and forward models, it may be possible to accurately localize these sources. This sets the limit of what linear analysis can, or cannot, do.

2. Introduction

The term "virtual electrode" refers to the effect of applying, to multi-channel data, a spatial filter designed to enhance a particular spatial position. Specifically, given a data matrix (time × channels) and a physical position, the method aims to produce a time series (time × 1) similar to what would have been obtained by placing a real electrode at that position. Examples are spline interpolation between electrodes (e.g. http://wiki.besa.de/index.php?title=BESA_Research_Manual), or beamforming (van Veen et al 1997). The term is justified by two properties: the filter targets a particular *spatial position*, and it produces a single *time series*, as would a real electrode placed at that position. Like a real electrode, the virtual electrode can be positioned anywhere within the brain, and the associated spatial filter mimics spatial attenuation with distance from the tip of a real electrode. The term is motivated also by 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).

The prospect of a clinical application is appealing and motivates efforts to enhance the performance of these 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. Simultaneous recording from each of I distinct positions (e.g. one in each of the ~120000 cubic millimeters of the human brain) might result in a data matrix of rank up to I. In contrast, recording from the same number of virtual electrodes cannot yield a data matrix of rank greater than J, the number of sensors or electrodes (e.g. 128 for a typical EEG system). Clearly, 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). The "image" at any point in time is constrained to be a function of at most J distinct spatial patterns across the imaging field, rather than the $I \gg J$ independent voxels or pixels of an optical imaging technique. Likewise, the time course of any pixel or voxel is constrained to be a linear combination of the J sensor signals, rather than independent from that of the $I - 1$ other voxels.

These limitations are, of course well known, as they reflect the fundamental indeterminacy of the inverse problem (Baillet 2010). However, this is often lumped together with issues of sensor calibration, head position tracking, noise, or properties of a specific method. These collectively affect the quality of the *approximation* of an ideal virtual electrode, as quantified by concepts such as *point spread*, *leakage* and *crosstalk*, captured by a *resolution matrix* in the case of imaging. This conflates fundamental limitations (that cannot be overcome), and method or equipment-based limitations, making it hard to partition between limitations amenable to progress and those that are not.

In summary, the metaphor of virtual electrode is useful to convey the abilities of magnetic or electric source 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.

3. Concepts, methods, and results

I first describe the theoretical background, taking 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, but I do not implement or test them.

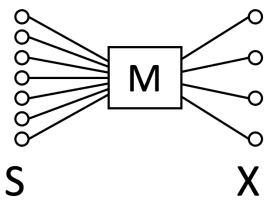
3.1. Data model

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. 1). 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: $I \gg J$, but the assumption $I = J$ or $I < J$ may be useful as an approximation. For example, a subset of sources with similar positions and activity might be assimilated to a single source, or some sources might be quiescent during a time interval during which an algorithm that requires $I \leq J$ can be applied.



$$x_j(t) = \sum_i m_{ij} s_i(t)$$

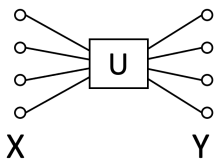
Fig. 1. Data model. The observations $\mathbf{X}$ (e.g. EEG signals) are related to brain sources $\mathbf{S}$ via a linear mixing process represented by $\mathbf{M}$. The signal of each of the J channels is the weighted sum of I brain source signals.

3.2. Analysis model

Typically, we are interested in (a) one or more columns of $\mathbf{S}$, representing the *time series* of activity of brain sources of interest, and (b) the corresponding columns of $\mathbf{M}$ insofar as they might allow to infer the *spatial positions* of those sources. These goals are distinct, although sometimes pursued together (as in magnetic source imaging and beamforming). Of course, neither $\mathbf{S}$ nor $\mathbf{M}$ are observable, the goal of analysis is to infer them or their properties from the data.

3.2.1 Linear analysis

Since the sources were mixed linearly, it makes sense to try to recover them using linear methods. Typically, an analysis matrix $\mathbf{U}$ of size $J \times K$ is applied to the data to yield a matrix $\mathbf{Y}$ of size $T \times K$ of transformed data (Fig. 2).



$$y_k(t) = \sum_j m_{jk} x_j(t)$$

Fig. 2. Linear analysis. Component waveforms $\mathbf{Y}$ are the product of observations $\mathbf{X}$ by an analysis matrix $\mathbf{U}$. Each component signal is the weighted sum of all observation channels.

Each column of $\mathbf{Y}$ is referred to here as a *component*. The number K of columns can be smaller or larger than J, however if it is larger the components are linearly dependent and thus correlated: the rank of $\mathbf{Y}$ cannot be greater than that of $\mathbf{X}$. If $K = 1$, the single-column matrix $\mathbf{U}$ is a *spatial filter* (see below).

The matrix $\mathbf{U}$ can be can be predefined, as in re-referencing, Laplacian, or current source density (CSD) transforms, or it can be derived based on 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 which 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 useful or informative. For example, a column of **Y** might approximate a column of **S**, or a compound pattern with a functional meaning such as an event-related potential (ERP).

3.2.2. Vector spaces

It is helpful to reason in terms of vector spaces. The signal $x_j(t)$ picked up by an electrode or sensor is a linear combination of brain source signals $s_i(t)$, and thus it belongs to the *vector space* spanned by those signals. That space is of dimension I at most, although it might be smaller if the $s_i(t)$ were linearly dependent.

The J columns of the observation **X** span a *subspace* of the source space, of dimension of at most J. The mixing matrix **M** projects onto this subspace (Fig. 3), and thus a given **X** might reflect many different patterns of brain activity.

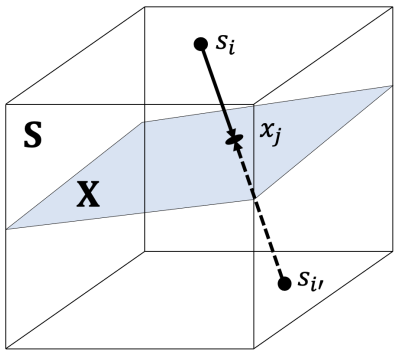


Fig. 3. The space spanned by the columns of the observation matrix **X** is a subspace of the space spanned by the brain signals **S**. Brain source signals are projected onto this subspace. Different sets of source signals may project on to the same observation: measurement techniques such as EEG cannot distinguish between them.

It commonly occurs that the dimension of the space spanned by **X** is smaller than J (**X** is rank deficient), for example if the data are re-referenced (EEG) or processed by a noise reduction algorithm such as SSS (signal space separation) in MEG.

From the data **X** is linearly derived the component matrix **Y**, which thus spans a subspace of the space spanned by **X**, of dimension at most J or K, whichever is smaller. For some analysis methods (e.g. PCA) the columns of **Y** are mutually uncorrelated. In that case, the subspace spanned by a subset of columns is orthogonal to the subspace spanned by the other columns. Each vector of one is uncorrelated with every vector of the other.

3.2.3. Spatial filter

If $K = 1$, the analysis matrix consists of a vector **u** of weights to be applied to channels of **X** before adding them. This is known as a *spatial filter*. Here, "spatial" refers to the J-

dimensional space of measurements, although it can sometimes be understood also in the sense of selectivity in the physical space of brain sources. The term "virtual electrode" evokes selectivity in that sense.

A linear combination of spatial filters is a spatial filter, and thus filters form a vector space, of dimension J because each filter is defined by J coefficients. A nice property is that, whatever the source $\mathbf{s_i}$, there exists a spatial filter $\mathbf{u}_k$ that can suppress it. A trivial example is a filter with coefficients that are all zero except for two channels j and j' such that $u_{jk}/u_{j'k} = -\,m_{ij}/\,m_{ij'}$. Additional useful properties are:

- From dimensionality arguments (Fig. 3), we know that there exist other sources with the same ratio $m_{ij}/\,m_{ij'}$. From the point of view of this filter, all those sources are equivalent and are suppressed. I call this the "null set" of the filter (see Fig. 6B for an illustration).
- In addition to this filter, there may exist other filters $\mathbf{u}_{k'}$, distinct from $\mathbf{u}_k$, that suppress $\mathbf{s}_i$. Together, these filters form a subspace of dimension $J-1$ of the space of filters.
- A spatial filter $\mathbf{u}$ can suppress up to $J-1$ sources (or sets of equivalent sources), while still transmitting some information.
- Sources other than those few are *not* suppressed. The filter remains sensitive to most sources in the brain.

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 reflecting the compound activity of all the other sources, those that were not silenced. Those other sources may be numerous and distributed across the brain: a spatial filter cannot be perfectly spatially selective, contrary to a real electrode.

3.2.4. Denoising

A single spatial filter that suppresses $\mathbf{s}_i$ produces a single time series, a very impoverished version of the data $\mathbf{X}$. However, it was noted that there exists a subspace of dimension J-1 of filters that do the same job. Together, the outputs of these filters provide a much richer picture of the data. Based on them, we can reconstruct a "denoised" version of the data.

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 be silenced by multiplying those components by zero and applying the inverse of $\mathbf{U}$. A "denoising matrix" can thus be defined as

$$\mathbf{D} = \mathbf{UEU^{-1}}$$

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:

$$\widehat{\mathbf{X}} = \mathbf{XD}$$

as illustrated in Fig. 4. The matrix $\widehat{\mathbf{X}}$ has the same dimensions as the original data and can be substituted for it in further analyses, with the caveat that it is not of full rank. Furthermore, if a forward model is used (see next section), that model needs to be adjusted to take into account the non-uniform gain due to the denoising matrix.

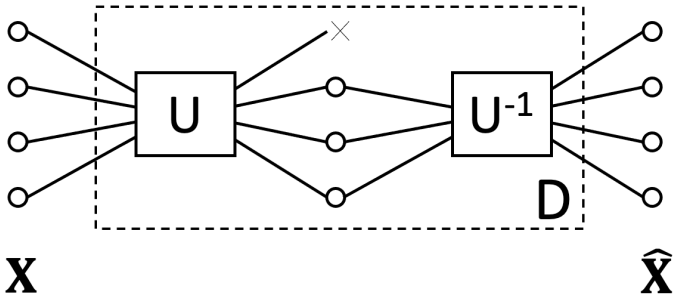


Fig. 4. Denoising. If the observation matrix is contaminated by a noise source that can be isolated by an analysis matrix **U**, zeroing that component and applying $\mathbf{U}^{-1}$ will produce a "clean" observation matrix that is invariant with respect to that noise.

3.2.5. Virtual electrode

A real electrode is sensitive to sources near its tip, and insensitive to distant sources (Fig. 5 left). To reproduce that property, a spatial filter would need to put a null on each of those distant sources (Fig. 5 right). In other words, most coefficients of $\mathbf{Mu}$ should be zero. As these rows number in billions (I), whereas available parameters 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.

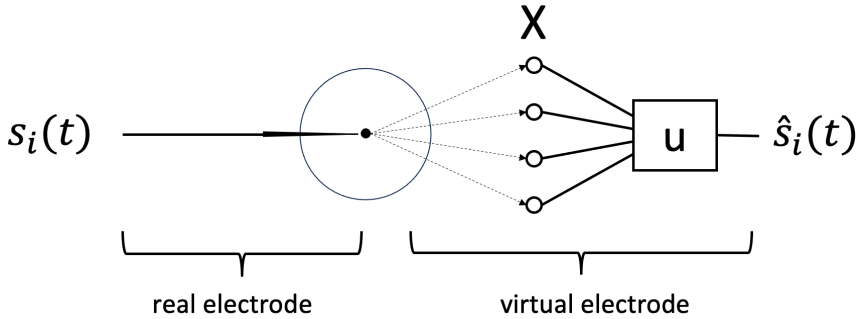


Fig. 5. A real electrode (left) is selectively sensitive to brain sources near its tip and blind to distant sources. A virtual electrode (right) attempts to reproduce this property by applying a spatial filter to the observation matrix **X**.

3.2.6. Data-driven analysis methods

So far, we argued for the existence of analysis matrices **U** or spatial filters **u,** but we did not say how they are found. In practice, the analysis matrix **U** may be predetermined, or calculated by applying a data-driven 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 **Y** are mutually uncorrelated. Second, their variance adds up to that of **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.

Interestingly, the *last* principal component is the linear combination of columns of **X** with the *least* possible amount of variance. The last column of **U** thus defines the spatial filter that best suppresses all sources present. If only $J-1$ sources were active, that spatial filter would cancel 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 $\mathbf{X}$ that are *maximally correlated* with another dataset $\mathbf{X}'$. Joint decorrelation (JD) seeks to maximize the *ratio of variance* between two subsets of the data, or between the data and a filtered version of the data. These methods share with PCA the property that columns of $\mathbf{Y}$ are mutually uncorrelated.

An individual column of the transformed data matrix $\mathbf{Y}$ can be understood in two ways. Either as *selecting* an individual source or group of sources, or as *excluding* one or more such sources (up to $J-1$). Classic accounts focus on the former, this paper focuses on the latter.

3.2.7. Existence vs identifiability

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

3.2.8. Localization and imaging

The previous discussion made no reference to the physics or geometry of the problem. The linear analysis methods are agnostic to those aspects, and in return they have nothing to say about the mixing matrix $\mathbf{M}$ or 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, and a *forward model* that specifies the source-to-sensor gain coefficients (row of $\mathbf{M}$) 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, we propose an additional approach.

Given a spatial filter that successfully cancels a source of interest, the source model and forward model can be used to find the possible source positions. The result of this search is referred to here as the *null set* of the spatial filter. The null set is typically *spatially extended* (multiple sources are equivalent because of the reduced rank of $\mathbf{X}$, in particular those for which $\mathbf{Xu}$ is zero), but also *sparse* within the search space allowed by the source model (examples are given below). The search process can also be cast as inference, the source model serving as a prior.

The null set constrains the position of the source, but it does not in general narrow it to a point. However, if multiple filters are available that all suppress a source, its location can be narrowed to the *intersection* of their null sets, which in some cases might be reduced to a point. In particular, supposing that a BSS method has successfully isolated the desired source within a column j of $\mathbf{Y}$ uncorrelated with other columns, the other columns $j' \neq j$ of $\mathbf{U}$ constitute spatial filters that suppress that source. In that case the source can be located as the intersection of $J-1$ null sets, a strong constraint.

The source model permitting, these solutions can be used to derive a visual rendering of inferred brain activity that can be related to anatomical knowledge. If the source model includes (or favors) spatially-extended sources constrained by anatomical knowledge, the image might appear realistic. However, one cannot use it as experimental *support* for details of that image. Even if the image gives the illusion of "seeing" a large number of individual pixels or voxels, the patterns that they form are assembled from a limited basis of J patterns, and the time course of each pixel/voxel is assembled from a basis of only J components.

3.3. Simulation within a toy world

The previous concepts are abstract. In particular it may be hard to see a relation between the signal spaces and the physical space of source and sensor locations (3-D for spatial position, 5-D if we include source or sensor orientation, a higher dimension if we include shape parameters, etc.). This section introduces a setup with simplified geometry to help clarify the geometric versus algebraic aspects of mixing and analysis. Sources and sensors are within a plane, and source-to-sensor propagation depends only on distance between source and sensor. The geometry of sources and sensors is thus fully specified by their coordinates within this 2-D space.

Specifically, a set of sensors is positioned along the margin of a disc (Fig. 1A, blue dots), and one or more sources are placed 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.

3.3.1. One source, two sensors

Two sensors can be combined linearly, $y(t) = u_j x_j(t) + u_{j'} x_{j'}(t)$ to form a spatial filter with a gain that depends on the spatial position of the source, as illustrated in Fig. 1B. Here, the coefficients were chosen so that the gain is zero for a source at the position of the green cross. Interestingly, the gain is also zero along a dark line, implying that other sources situated along that line would be suppressed too. This is referred to as the "null set" of this filter.

This illustrates three of the properties mentioned earlier. First, for any source, there exists a spatial filter that can suppress it. Second, that filter generally suppresses *multiple sources* (the null set), that here appear to fall along a curve in 2-D space. Third, the filter *is sensitive* to sources situated in most other locations. From the curvilinear aspect of the null set in Fig. 6B it is tempting to assume that it constitutes a manifold of the 2-D space of locations. However, the set can take more complex shapes (e.g. Fig. 8B) that are not consistent with a manifold.

Figure 6C shows a similar plot based on a different pair of sensors. Again, parameters are tuned so that gain is zero for a source at the green cross, and again there is an extended null set, distinct from the first.

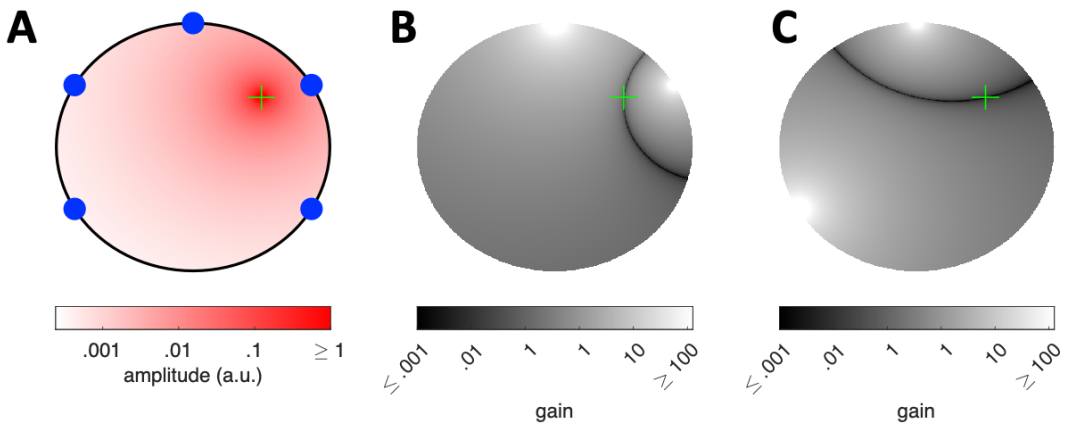


Fig. 6. A: simplified setup consisting of an array of sensors (blue dots) placed at the periphery of a disc. Sources within the disc radiate within the plane with an amplitude that decreases with distance as $1/r^2$ (pink color scale). B: gain as a function of source position within the plane for a linear combination of two sensors (coefficients 1 and -0.68) coded as a gray scale. C: same for a different pair of sensors (coefficients 1 and 0.075). The dark lines trace locations of zero gain ("null set" of the filter).

Filter weights might be obtained analytically in this simple configuration, but an easier method, applicable to more complex configurations, is to simulate the activity of the source(s) (e.g. using Gaussian noise), derive a simulated sensor data matrix using a forward model, and apply PCA to that matrix. If only one source is active, the data matrix is necessarily of rank one, and thus the last row of the PCA matrix maps the data to zero, i.e. the last column of the PCA matrix defines the desired filter.

This data-driven approach can be also used to discover the *location* of a source if that source is the only one active. Applying PCA to a pair of 2 sensors, the last column of the PCA matrix gives the coefficients of a filter that suppresses the source, from which infer a first null set (dark line in Fig. 6B). With a second sensor pair we obtain a second null set, which narrows the source position to its intersection with the first (a point in Fig. 6C). In this simple 2-D world, the location of a source can be determined with *arbitrary precision* from data recorded with just three sensors.

More generally, if $K > 2$, there are multiple spatial filters that can suppress a single source. With one active source, the K-channel sensor data matrix is of rank one, and thus columns 2 to K of the PCA transform matrix all constitute null filters. Since any linear combination of such filters is also a null filter, these columns form the basis of a space of filters of dimension $K - 1$, subspace of the space of K-channel spatial filters. To each of these $K - 1$ filters is associated a null set, and the location of the source is constrained to their intersection.

Figure 7A-D shows four such filters obtained by applying PCA to the 5-channel data matrix in response to a single source of Gaussian noise. The zero-gain lines take on diverse shapes (variable according to the sample of noise) that all pass by the source. Taking the max over gain maps produces one dark spot from which the source position can be accurately determined (E).

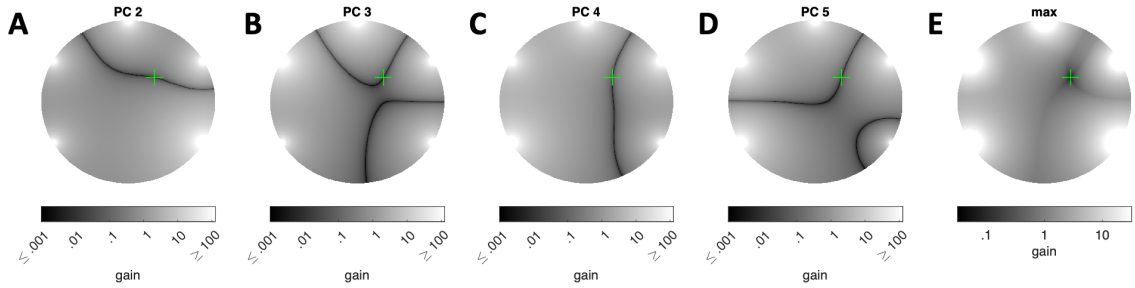


Fig. 7. A-D: gain maps for PCs 1 to 4 from a PCA applied to data from 5 sensors in response to a single source located at the green cross. The source waveform was Gaussian noise. E: max over gain plots. The dark spot indicates the position of the source.

3.3.2. Multiple sources

What if there are more than one source? If $1 < I < J$, the matrix of sensor data is of reduced rank, and thus the last column of the PCA matrix maps the data to zero. Thus, it defines a spatial filter that suppresses *all* I sources. With 5 sensors, up to 4 active sources can be suppressed in this way, as illustrated in Fig. 8A. The spatial filter that suppresses all four sources is unique, to a multiplicative constant, and the same would be true for other numbers of sensors/sources as long as $I = J - 1$ (e.g. Fig. 8B). With $I = J - 2$ sources, the last *two* columns of the PCA matrix map to zero (Fig. 8C, D), thus spanning a *two-dimensional subspace* of spatial filters that can suppress these $J - 2$ sources.

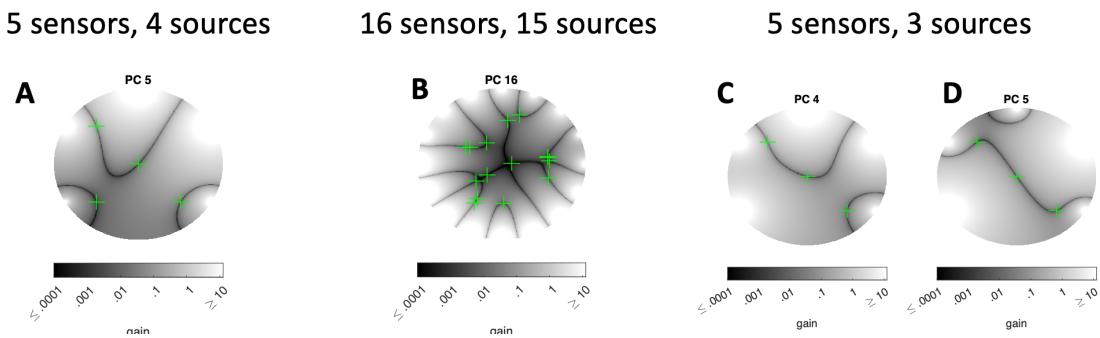


Fig. 8. A: gain pattern for the unique (to a multiplicative constant) filter that can suppress 4 sources given 5 sensors. B: same, with 15 sources and 16 sensors. C and D: gain patterns for two spatial filters that that can suppress 3 sources given 5 sensors. These form a basis of the 2-dimensional subspace of such filters. All these filters were derived using PCA.

Here again, these methods can be used to help discover the locations of the sources.
The location of the source(s) is in some cases ambiguous (e.g. Fig. 8A and B), in others less so (Fig. 8C, D) but there appears to be no fundamental limit on the *precision* of the estimates. Precision is limited mainly by the resolution of the search grid used to locate the null set within physical space, if not by numerical or physical precision. There is no equivalent of "point spread" as invoked to quantify localization uncertainty in MSI (e.g. Hauk et al 2022).

To recapitulate, given J sensors, a spatial filter *exists* that can suppress up to $I = J - 1$ sources while remaining sensitive to sources in other positions. That spatial filter *can be found* using PCA if those sources are the only ones that are active. Those sources can then be *located* (possibly with some ambiguity) as belonging to the null set of that filter, given a source model and a forward model. For a smaller number of sources ($I \leq J - 2$), filters that

suppress the I sources form a space of dimension $J - I$. If those are the only sources that are active, a basis of this subspace can be found as the last columns of the PCA transform matrix. The sources may then be located more precisely by taking the intersection of their null sets.

3.3.3. As many sources as sensors, or more

In practice, there are usually many more sources than sensors. The brain contains billions of neurons, each of which might contribute to the measurement, to say nothing of physiological, environmental, or instrumental sources of noise. In particular, if each sensor contributes instrumental noise, there must always be *at least* J active sources. The previous account may nonetheless remain useful as an approximation.

The case $I = J$ is tantalizing because we know that there exists a filter (or set of filters) that can suppress any subset of those sources, but we are unable to find those filters, at least using the PCA-based method. The case $I = J$, while presumably uncommon, is the basis for many BSS methods such as ICA.

If the activity of the I sources is *non-stationary* (as is often the case), it might happen that the number of non-quiescent sources $I' < J$ over a subset of the time axis large enough for PCA to be effective. One can then find a filter (or set of filters) to suppress the I' non-quiescent sources, and possibly locate them. Non-stationarity can also be exploited using Joint Decorrelation (JD, also known as Denoising Source Separation, DSS, or Common Spatial Patterns, CSP, see review in de Cheveigné and Parra 2014). In a nutshell, the covariance matrix of the data is calculated over a subset of the time axis, and diagonalized jointly with the covariance matrix calculated over the entire time axis. The first (respectively last) column of the transform matrix yields the component with the highest (respectively lowest) ratio of power over the subset relative to the entire data. Non-stationarity is one of three cues that can be exploited by BSS algorithms (e.g. Anemüller 2007), the other two being spectral mismatch and non-gaussianity (Cardoso 2001).

BSS, if successful, yields an analysis matrix that is the inverse of the unknown mixing matrix, $U = M^{-1}$, to a permutation. This is the holy grail of brain data analysis, because the investigator can directly observe the activity of individual brain sources, although there might be uncertainty as to which observation matches which source due to the permutation.

Knowledge of the unmixing matrix aids source localization because the source that maps to component j is located within the intersection of the null sets of all columns $j' \neq j$, as illustrated in Fig. 9 for the 5-sensor toy world configuration.

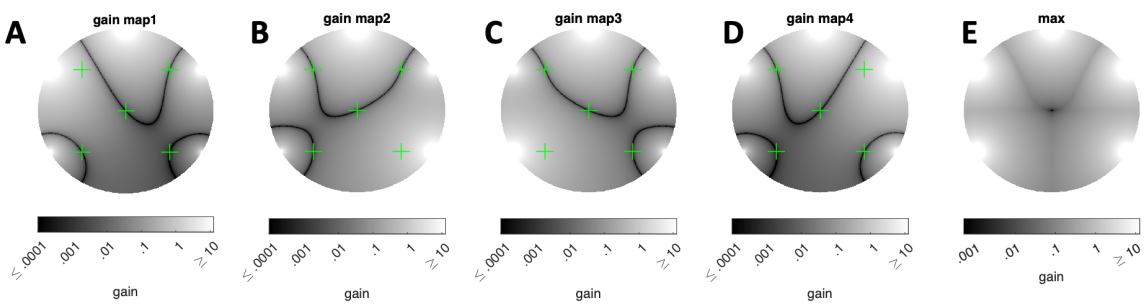


Fig. 9. A-D: gain maps for columns 1-4 of the unmixing matrix $U = M^{-1}$ associated with the 5 sources indicated as green crosses. For each gain map, the null set (black line) includes all but one of the source locations, different for each column. The 5th source (center) is located within the null sets

of all four maps, and its precise location can be found as the position of the minimum of the maximum over these gain maps (E). The difference with Fig. 7 is that here there are 4 active sources rather than one.

This result holds even if the sources that map to components $j' \neq j$ are not well separated, i.e. in the plausible situation where the analysis can only reliably isolate a single source, the others being lumped into a set of $J - 1$ components. It also holds even if the target source is not perfectly isolated within the component that it maps to, as long as its contribution to the other components is zero. Again, precision is limited mainly by the resolution of the search grid used to delineate the null sets (500×500 in these simulations).

To recapitulate, localization of a source involves two steps. In the first step, a data-dependent method is used to estimate one or several spatial filters that suppress that source. In the second step, a source model and a forward model are used to search for the null sets of each of these filters, and the location is derived from their intersection. The "source model", in this simulation, is the assumption of point sources within a disc. Source and forward models intervene only in the second step, which is applicable to any set of spatial filters, regardless of how they were obtained.

3.3.4. Beamformer

It is interesting to see how this relates to the result of applying a linear constrained minimum variance (LCMV) beamformer, a popular method to isolate an individual source or locate it. A LCMV beamformer algorithm produces a spatial filter that (a) passes activity from a source at 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 10A gives an example of applying the algorithm to a 5-source, 5-sensor configuration, with the beamformer targeting the upper right source (red circle). The resulting spatial filter perfectly suppresses all other sources while preserving the target source. The gain map is identical to those plotted in Figs. 8A and 9D, consistent with the earlier argument that a filter suppressing 4 sources within data from 5 sensors is unique.

Several features of the LCMV solution are not consistent with the idea of a virtual electrode. First, based on dimensionality considerations, it was argued that an extended set of sources may be equivalent from the point of view of a spatial filter. Thus, the beamformer solution derived for one location is *exactly* the same as that derived from other locations with in this set. Regardless of this argument, it is clear from Fig. 10A that the gain is non-zero for most locations within the disc, contrary to what would be expected of an electrode.

The beamformer solution is also somewhat fragile. The solution illustrated in Fig. 10A was obtained with data from sources that were *perfectly decorrelated* by orthogonalizing the matrix generated by the Gaussian noise process used to simulate source activity. Omitting that step yields the solution illustrated in Fig. 10B. The gain appears to be zero for one interfering source (upper left) but not the others, implying that the beamformer output is *not invariant* with respect the activity of those sources.

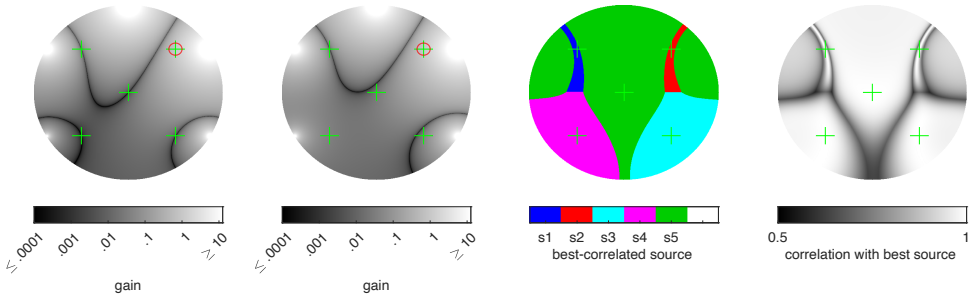


Fig. 10. A: gain map of a beamformer tuned to target one source among 5 active sources (green crosses). The targeted source is indicated by a circle. B: same, but the source time series of are slightly correlated between sources ($|\rho| \leq 0.04$), resulting in an imperfect solution. C: identity of the source that is best correlated with the beamformer output as a function of the target position. Color-coded sources are numbered clockwise starting from the upper left (s1) and ending with the center (s5). D: correlation between the beamformer output and the best correlated source.

The LCMV beamformer has been proposed as a means to locate a source within the brain, by scanning physical space and monitoring the output, for example for some property indicating that the beamformer has latched on to the desired source (e.g. spectral content, or ERP). Figure 10CD shows the result of such a scan. Panel C shows the identity of the source most highly correlated with the beamformer output. The search space is, as expected, partitioned into zones that appear to include the respective sources, providing a rough basis for localization. Panel D shows the correlation of the beamformer output with the best-correlated source.

Panel D could be used as a basis to refine the position estimates by searching for the maxima of correlation, but it appears that patterns are rather wide. Sources 1 and 2 are constrained to two narrow "horns", but the patterns associated with the other sources are more extended, and that associated with source 5, center, even includes two lateral "echos". From these patterns, one might be tempted to conclude that the sources themselves are spatially distributed, whereas we know by construction that they are not.

3.4. Extrapolating to MEG and EEG

Application to real MEG or EEG is beyond the scope of this short paper. While it is of interest to know whether the methods work, the many details and constraints to discuss would lead to a much longer paper, and the numerous details might obscure the points made here. Here, I outline a few issues to be addressed.

A first issue is that there are many more sources than sensors, so any result that requires $I < J$ or $I \leq J$ can only be useful as an approximation. In particular, sensor noise implies at least J sources, so the issue of insufficient observation dimensionality cannot be tackled by merely increasing the number of sensors (although doing so might improve the approximation).

A second issue is the complexity of the source model. A dipolar source constrained to the cortical surface and perpendicular to that surface has the same dimensionality as our flat disc, but allowing the dipole to be located anywhere in the brain, with any orientation, implies a 5-D parameter model, and augmenting it to allow for extended sources, or source pairs, would further complexify the model. There is no obvious reason why the present results cannot be extended to such a source model, but experience with high-dimensional spaces teaches us to be wary. The shape of a null set is certainly hard to apprehend, and it is not clear (at this point) how many null sets are needed to fully localize a source. The cost of search is also a concern, although if source and sensor models are continuous and differentiable (as here) efficient search strategies should be possible.

A third issue is the complexity and accuracy of the forward model, particularly for EEG, as it depends not only on sensor calibration and head tracking, but also on a realistic model of the complex conductivity patterns within the head.

4. Discussion

This paper offers an account that is original in its emphasis on spatial filtering as a means to *suppress*, rather than select, a specific source of brain activity. A feature of this account is that it attempts to separate clearly algebraic aspects (existence of cancellation filters, their vector-space structure and dimensionality), data-driven estimation (PCA, ICA or other algorithms), and issues related to source models and forward models relevant for source localization or imaging. Two goals may be achieved using this approach: suppression of activity specific to one or a few sources, so as to obtain a picture of other activity invariant to those sources, and accurate localization of those sources.

4.1. Summary of results

The main results are as follows (with no claim that they are novel or non-trivial), divided into algebraic, data-driven, and source/forward model-related.

1) Algebraic properties:
   a) With J sensors, the set of all spatial filters forms a vector-space of dimension J. Spatial filters that can suppress a source $\mathbf{s_i}$ form a subspace of dimension $J-1$. Filters that simultaneously suppress a set of I' sources form a subspace of dimension $J-I'$.
   b) Spatial filters can be used to make the observed data invariant with respect to activity of one or more sources (denoising). If I' sources are suppressed, the denoised data matrix $\widehat{\mathbf{X}}$ is of rank $J-I'$ at most.
   c) To each filter corresponds an extended set of sources that that filter suppresses (null set).
2) Data-driven estimation
   a) The required filters and analysis matrices can sometimes be obtained with data-driven methods. If $I' < J$ sources are active, the space of filters that suppress them is spanned by the last $J-I'$ columns of the PCA transform matrix. Data-driven estimation requires that source activities are not linearly dependent, and some methods might also require them to be uncorrelated.
   b) if $I' = J$ sources are active, and a BSS method such as ICA is successful, the $J-1$ filters that suppress source $\mathbf{s}_i$ are found as the columns that isolate the $J-1$ other sources.

c) More generally, if an analysis method succeeds in isolating source $\mathbf{s}_i$, the other columns of the transform matrix span the subspace of filters that suppress $\mathbf{s}_i$. It is not necessary that $\mathbf{s}_i$ be perfectly isolated, nor is it necessary that other sources be demixed in the output of the analysis.

3) Localization
   a) Given a spatial filter that suppresses $\mathbf{s}_i$, together with a source model indicating possible source positions, and a forward model that specifies source-to-sensor gains, $\mathbf{s}_i$ is known to be located in the null set of the filter.
   b) If several spatial filters are available, a null space can be derived from each, and $\mathbf{s}_i$ is known to be located within their intersection.
   c) The source model can, in principle, be extended to include various shapes (e.g. extended sources) and other characteristics or constraints, such as anatomical limits, multiple sources (e.g. one in each hemisphere), etc.. In that case, it may be possible to perform rudimentary imaging, with the understanding that aspects of the image may reflect the source model more than the data.
   d) This approach to localization can be applied using spatial filters obtained with any method.

Quid the "virtual electrode"?

None of these properties seems to fit the concept of virtual electrode. Judging from Figs. 6-10, spatial filters tend to have patterns of high-sensitivity (the electrode "tip") that are broad and convoluted. Patterns of low-sensitivity (null sets) are narrower, and while convoluted for individual filters, their intersection may be more narrowly localized. This property better fits the metaphor of "virtual scalpel", by which activity may be silenced (and by the same token localized) by a spatially precise intervention.

To be fair, methods might be found that produce a spatial filter with a more favorable pattern. However, the space of spatial filters is rather small (dimension at most J) and inhabited with solutions produced by the various other methods we considered (PCA, ICA, etc.). As an example, a beamformer solution that cancels $J-1$ sources is necessarily identical to the last row of the PCA matrix obtained when those sources are the only one present (compare Figs. 8A, 9D and 10A): no method can do a better job. The concepts described here, encapsulated by the metaphor of "virtual scalpel", seem a better approach to understanding what can be achieved with linear analysis.

Future directions

A first, obvious direction is to test these ideas with real EEG and MEG. This could usefully be preceded by simulations with a more complex "toy world", for example with higher source space dimensionality, and by simulations with a synthetic source (ground truth) embedded within real EEG/MEG data.

A second direction is to work out the relation between these and other methods (Baillet et al 2010, Hauk et al 2022). There are likely to be deep connections, grounded in the algebra (e.g. Mosher et al 2003), but these may be obscured by details relative to anatomy, instrumentation, noise, and so-on. The present approach is more circumspect than others with respect to "virtual electrodes" and imaging, and more upbeat in regard to the ability to localize sources with precision. These discrepancies should be resolved.

A third direction is to further explore the capabilities of the method. With J sensors at most J sources can be resolved, but it might be possible to exploit non-stationarity to push this limit. For example, it might be possible to focus analysis on the initial portion of an ERP, derive a set of spatial filters that suppress it, and thus localize its source accurately. The operation could be renewed to trace later portions of the ERP. A similar approach might work for ictal seizures. To what extent this is workable needs to be explored.

Conclusions

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, and with appropriate source and forward models, it may be possible to accurately localize these sources. This sets the limit of what linear analysis can, or cannot, do.

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

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

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

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

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

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