Volume Conduction Influences Scalp-Based Connectivity Estimates

Clemens Brunner [1,2]*, Martin Billinger [3], Martin Seeber [2,4], Timothy R. Mullen [5] and Scott Makeig [6]

[1] Institute of Psychology, University of Graz, Graz, Austria, [2] BioTechMed-Graz, Graz, Austria, [3] Department of Otolaryngology, Hannover Medical School, Hannover, Germany, [4] Institute of Neural Engineering, Graz University of Technology, Graz, Austria, [5] Qusp Labs, San Diego, CA, USA, [6] Swartz Center for Computational Neuroscience, University of California, San Diego, San Diego, CA, USA

INTRODUCTION

Electroencephalographic (EEG) signals recorded from the scalp surface are generally highly correlated. Each channel is a linear mixture of concurrently active brain and non-brain electrical sources whose activities are volume conducted to the scalp electrodes with broadly overlapping patterns (Nunez et al., 1997). This property is particularly relevant to connectivity analyses, which seek to detect and characterize active interactions between brain regions. Therefore, meaningful connectivity patterns can be derived only from measures of cortical source activities and not directly from EEG channel activities (Michel and Murray, 2012). However, this poses a serious problem, since estimating the nature, number, brain (or non-brain) locations, and time courses of the active sources contributing to the scalp EEG is not straightforward (Baillet et al., 2001).

Several methods have been proposed to estimate source activities from multi-channel EEG recordings, thereby removing the confounding effects of volume conduction. These methods can be grouped into three categories: (a) simple spatial filters that seek to reduce correlations between scalp channels based on idealized assumptions; (b) more complex spatial filters that seek to estimate net activities within ROIs based on detailed neurophysiological head models; and (c) blind spatial source separation methods that seek to separately identify source signals by exploiting source signal information differences. Whereas simple spatial filters such as bipolar derivations and Laplacian filters can reduce, to some extent, correlations among scalp-recorded channels (Fisch, 2012), more sophisticated spatial filtering methods use inverse imaging methods to estimate the time courses of cortical sources in given or estimated region of interest (ROI) (Baillet et al., 2001). Blind source separation techniques, in particular independent component analysis, by contrast, learn spatial filters from the EEG time courses that separate the data into constituent independent source activities. Their corresponding brain (or non-brain) locations can then be estimated using neurophysiological inverse imaging methods (Makeig et al., 1996; Jung et al., 2001; Delorme et al., 2012).

Vector autoregressive (VAR) models are versatile tools for analyzing multivariate time series, including multi-channel EEG or multivariate source activities. VAR models predict current values of time series from their recent past (Lütkepohl, 2005). Importantly, they can be used to derive various electrophysiological connectivity measures (Schlögl and Supp, 2006). Volume conduction in biological tissue can be modeled as instantaneous propagation of activity from sources to recording channels. The resulting zero-phase connectivity may be treated as noise added to lagged connectivity patterns of interest. Although some measures, including the imaginary part of the coherency (Nolte et al., 2004), are insensitive to zero-phase connectivity, measures derived from VAR model coefficients do not include such zero-phase terms. Thus, volume conduction effects are

not accounted for by the model and affect the correlation structure of the model residuals, which are normally assumed to be uncorrelated.

Popular connectivity measures derived from VAR models include the Directed Transfer Function (DTF) (Kaminski and Blinowska, 1991) and the Partial Directed Coherence (PDC) (Baccalá and Sameshima, 2001). Whereas the PDC is defined in terms of the system matrix (a frequency domain representation of the VAR model), the DTF is based on the inverse of the system matrix. Viewing lagged dependencies between source signals as information flow, the DTF may be said to be normalized by the inflow of information to some sink, while the PDC is normalized by the outflow of information from some source.

As we will demonstrate below, both the DTF and the PDC are indeed adversely affected by volume conduction from multiple sources to the scalp electrodes, in contrast to the claim of Kaminski and Blinowska in their recent opinion article (Kaminski and Blinowska, 2014). Thus, in general direct application of connectivity measures to scalp EEG signals produces less than accurate results and also does not allow their clear interpretation in terms of underlying source dynamics.

DTF IS SENSITIVE TO VOLUME CONDUCTION

Kaminski and Blinowska (2014) used an example to demonstrate that the DTF is not affected by volume conduction. In their example, they compared the DTF between all pairs of 18 scalp EEG channels to the DTF between the same channels overlaid with a sinusoid at 20 Hz. They interpreted the absence of a peak at 20 Hz in the DTF as evidence for their claim that the DTF is insensitive to volume conduction.

Upon closer inspection, this conclusion turns out to be based on questionable assumptions. First, the figure in their example clearly shows that the DTF obtained from the overlaid sinusoid are not identical to the original DTF. Not only do most of the connectivity measures exhibit troughs at 20 Hz, but the morphologies of some channel pairs are markedly different from the case in which no sinusoid is present. Second, the observed changes in the DTF can be explained by the fact that adding a sinusoid to all EEG channels with equal strength does not reflect volume conduction of a realistic brain source. If anything, their simulation only demonstrates changes in the channel-wise DTF when a single unrelated source signal (here artifactual and non-brain) is present on all channels. Further, their example implicitly confounds scalp and source signals. As described above, volume conduction is more accurately modeled as a mixing process—that is, producing channel activities that are linear combinations of all source activities. Our examples in the following sections will use this model.

ANALYTIC DERIVATION

We will now demonstrate that non-connected cortical sources generally exhibit spurious connectivity when measured between EEG channels. In general, a VAR model of order p can be written as

$$\mathbf{s}(t) = \sum_{k=1}^{p} \mathbf{A}_k \mathbf{s}(t-k),$$

where $\mathbf{s}(t)$ denotes a multivariate time series, $\mathbf{A}_k$ are the model coefficient matrices, and p is the model order. For the sake of simplicity, we will use only two sources and a model order of $p = 1$ (but note that the result is valid for an arbitrary number of sources and choice of model order):

$$\mathbf{s}(t) = \mathbf{A}_1 \mathbf{s}(t-1).$$

We begin by setting both off-diagonal elements in the model coefficient matrix to zero, imposing zero causal interaction between the two sources:

$$\mathbf{A}_1 = \begin{pmatrix} a_{11} & 0 \\ 0 & a_{22} \end{pmatrix}.$$

Transforming the model to the frequency domain via the Fourier transform yields the system matrix

$$\mathbf{A}(\omega) = \begin{pmatrix} 1 - a_{11}e^{-j\omega} & 0 \\ 0 & 1 - a_{22}e^{-j\omega} \end{pmatrix},$$

where ω is the angular frequency and $j^2 = -1$ is the imaginary unit.

The DTF is based on the transfer matrix $\mathbf{H}$ of the VAR model, which is the inverse of the system matrix $\mathbf{A}$. In our specific example, $\mathbf{H}$ is therefore also a diagonal matrix:

$$\mathbf{H}(\omega) = \mathbf{A}^{-1}(\omega) = \begin{pmatrix} \frac{1}{1-a_{11}e^{-j\omega}} & 0 \\ 0 & \frac{1}{1-a_{22}e^{-j\omega}} \end{pmatrix}$$

The DTF is just a normalization of the transfer function, which means that it is also a diagonal matrix. It follows that the DTF is zero for both off-diagonal elements, correctly indicating two non-connected sources.

Now we introduce volume conduction via a generic mixing matrix

$$\mathbf{M} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix}.$$

This mixing matrix relates the sources $\mathbf{s}(t)$ to the observed EEG signals $\mathbf{x}(t)$ such that

$$\mathbf{x}(t) = \mathbf{M}\mathbf{s}(t).$$

Substituting the VAR model into this equation, the mixed model coefficient matrix thus becomes

$$\tilde{\mathbf{A}}_1 = \mathbf{M}\mathbf{A}_1\mathbf{M}^{-1},$$

and the corresponding mixed system matrix

$$\tilde{\mathbf{A}}(\omega) = \mathbf{I} - \tilde{\mathbf{A}}_1 e^{-j\omega},$$

where $\mathbf{I}$ is the identity matrix. The mixed transfer function is the inverse of the mixed system matrix

$$\tilde{\mathbf{H}}(\omega) = \tilde{\mathbf{A}}^{-1}(\omega) = \begin{pmatrix} \tilde{h}_{11}(\omega) & \tilde{h}_{12}(\omega) \\ \tilde{h}_{21}(\omega) & \tilde{h}_{22}(\omega) \end{pmatrix}.$$

Plugging in the specific coefficients, the off-diagonal entries of the mixed transfer matrix can be written as

$$\tilde{h}_{12}(\omega) = \frac{m_{11}m_{12}e^{-j\omega}(a_{22} - a_{11})}{(a_{11}e^{-j\omega} - 1)(a_{22}e^{-j\omega} - 1)(m_{11}m_{22} - m_{12}m_{21})}$$

and

$$\tilde{h}_{21}(\omega) = \frac{m_{21}m_{22}e^{-j\omega}(a_{11} - a_{22})}{(a_{11}e^{-j\omega} - 1)(a_{22}e^{-j\omega} - 1)(m_{11}m_{22} - m_{12}m_{21})}.$$

Since the DTF is a specific normalization of the transfer matrix, its off-diagonal elements can only be zero if the corresponding off-diagonal elements of the transfer matrix are zero. In our example, both elements are only zero if $a_{11} = a_{22}$, that is, if both source activities are exactly equal. This specific case is highly unlikely to occur for real brain sources, which are separated by any distance and concomitant neural conduction delay. In general, the off-diagonal elements will thus be non-zero. This result demonstrates that given non-connected sources, spurious apparent connectivity is introduced into the result of VAR-based connectivity analysis when (as in actual data) volume conduction is present.

SIMULATION

Although the analytic derivation can be extended to more than two sources, computing and notating the individual elements of the mixed transfer matrix becomes quite unwieldy. Therefore, we simulated three cortical sources and used a realistic forward model of volume conduction from brain to scalp to numerically estimate their summed activity at three scalp EEG channels. For this simulation, we used the freely available Brainstorm toolbox (Tadel et al., 2011); our analyses are based on the ICBM152 template brain (Mazziotta et al., 1995) included in the toolbox. The source code for this analysis can be found in the Supplementary Material section. Once again, we start by specifying the ground truth, here the connectivity structure of the cortical sources consisting of one active connection from Source 1 to Source 3. The corresponding model coefficient matrix is

$$\mathbf{A}_1 = \begin{pmatrix} 0.9 & 0 & 0 \\ 0 & 0.5 & 0 \\ 1 & 0 & -0.5 \end{pmatrix}.$$

Next, we assume these three sources exhibit spatially coherent activity across upward-facing (gyral) cortical patches located just below electrodes placed at C3, C4, and F4 (see **Figure 1**). We compute the forward solution from the realistic template head model to obtain the mixing matrix $\mathbf{M}$ containing the weights that mix the activities of the three sources to the three scalp channels:

$$\mathbf{M} = \begin{pmatrix} 25.62 & 1.51 & -3.77 \\ 1.25 & 19.70 & 2.68 \\ 0.64 & 7.84 & 30.95 \end{pmatrix}$$
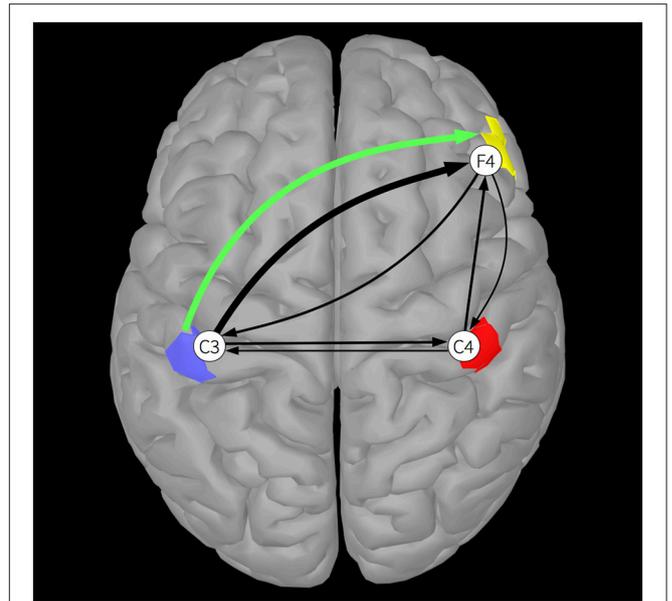
FIGURE 1 | A single DTF connection, averaged across all frequencies, between two of the three cortical sources (blue, red, and yellow patches), is shown by a green arrow. Black arrows indicate DTF connections estimated between EEG electrode channels C3, C4, and F4. The widths of the arrows are proportional to their DTF strength.

The matrix of mixed coefficients is

$$\tilde{\mathbf{A}}_1 = \mathbf{MA}_1\mathbf{M}^{-1} = \begin{pmatrix} 0.75 & -0.08 & 0.16 \\ 0.12 & 0.52 & -0.07 \\ 1.23 & 0.25 & -0.37 \end{pmatrix}.$$

The corresponding source DTF averaged over all frequencies is given by

$$\overline{\mathrm{DTF}}_{\mathrm{source}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0.66 & 0 & 0.70 \end{pmatrix},$$

whereas the corresponding channel DTF is given by

$$\overline{\mathrm{DTF}}_{\mathrm{channels}} = \begin{pmatrix} 0.98 & 0.08 & 0.16 \\ 0.14 & 0.98 & 0.08 \\ 0.72 & 0.17 & 0.62 \end{pmatrix}.$$

Figure 1 shows the results of the simulation. The green arrow indicates the ground truth connection between Source 1 (blue patch) and Source 3 (yellow patch). The black arrows visualize the estimated EEG channel connectivities. The width of the arrows is proportional to the strength of the DTF across all frequencies. Once again, it is apparent that volume conduction and summation of the projected signals at the scalp electrodes produces spurious connectivity results. Whereas there is only one ground truth source-to-source connection, non-zero values are produced between all channel pairs by applying DTF to the channel data. Note that spurious channel-to-channel connections are caused by volume conduction and not by indirect connections—similar connectivity patterns can also be

observed in PDC estimates of channel-to-channel connections. Here, the dominant estimated channel-to-channel connectivity is between simulated electrode channels placed directly above the coupled sources. In simulations in which the unknown source locations are more distant from the scalp electrodes, channel connectivity estimates may not suggest a particular source connectivity.

DISCUSSION

We have demonstrated both analytically and by numerical simulation that the DTF is influenced by volume conduction. In contrast to Kaminski and Blinowska (2014), we argue that connectivity measures should generally be calculated on cortical sources, because they are the actual sources of brain activity. Ideally, source activities can be obtained by separating each data channel signal into a sum of physically and physiologically distinct source processes whose interrelationships can also be modeled in terms of causal connectivity.

It is worth mentioning that we selected highly favorable EEG channel and source locations in our simulation example. Even in this best-case scenario, spurious connections arise between all channel pairs. However, most cortical sources do not necessarily project radially from cortical gyri to scalp locations just above them—with different source locations, we could have obtained a collection of suggested scalp channel connections far from

the originating sources, with no way to determine which source connections contributed to what extent.

In conclusion, while DTF and other connectivity estimators can be applied to either scalp channels or to (estimated) source signals, results are highly likely to be more accurate when the analysis is based on source activities. In addition, interpreting the results correctly in terms of regional brain interactions is also much more straightforward. This is because connectivity computed between EEG channels is heavily confounded by the broad volume conduction patterns associated with each regional EEG source.

AUTHOR CONTRIBUTIONS

CB led the article (which includes writing and contributing to all aspects of the article such as the analytic proof and the simulation example), MB laid out the analytic proof, MS contributed the simulation, TM provided input and comments/corrections on all aspects of the article, and SM provided input and comments as well as rewrote several paragraphs of the article.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fncom.2016.00121/full#supplementary-material

REFERENCES

Baccalá, L. A., and Sameshima, K. (2001). Partial directed coherence: a new concept in neural structure determination. *Biol. Cybern.* 84, 463–474. doi: 10.1007/PL00007990

Baillet, S., Mosher, J. C., and Leahy, R. M. (2001). Electromagnetic brain mapping. *IEEE Signal Process. Mag.* 18, 14–30. doi: 10.1109/79.962275

Delorme, A., Palmer, J., Onton, J., Oostenveld, R., and Makeig, S. (2012). Independent EEG sources are dipolar. *PLoS ONE* 7:e30135. doi: 10.1371/journal.pone.0030135

Fisch, B. J. (2012). "Polarity and field determinations," in *Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, eds D. L. Schomer and F. H. L. da Silva (Philadelphia, PA: Lippincott Williams & Wilkins), 143–151.

Jung, T.-P., Makeig, S., McKeown, M. J., Bell, A. J., Lee, T. W., and Sejnowski, T. J. (2001). Imaging brain dynamics using independent component analysis. *Proc. IEEE* 89, 1107–1122. doi: 10.1109/5.939827

Kaminski, M. J., and Blinowska, K. J. (1991). A new method of the description of the information flow in the brain structures. *Biol. Cybern.* 65, 203–210. doi: 10.1007/BF00198091

Kaminski, M. J., and Blinowska, K. J. (2014). Directed transfer function is not influenced by volume conduction—inexpedient pre-processing should be avoided. *Front. Comput. Neurosci.* 8:61. doi: 10.3389/fncom.2014.00061

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Berlin; Heidelberg: Springer-Verlag.

Makeig, S., Bell, A. J., Jung, T.-P., and Sejnowski, T. J. (1996). Independent component analysis of electroencephalographic data. *Adv. Neural Inform. Process. Syst.* 8, 145–151.

Mazziotta, J. C., Toga, A. W., Evans, A., Fox, P., and Lancaster, J. (1995). A probabilistic atlas of the human brain: theory and rationale for its development. *NeuroImage* 2, 89–101. doi: 10.1006/nimg.1995.1012

Michel, C. M., and Murray, M. M. (2012). Towards the utilization of EEG as a brain imaging tool. *NeuroImage* 61, 371–385. doi: 10.1016/j.neuroimage.2011.12.039

Nolte, G., Bai, O., Wheaton, L., Mari, Z., Vorbach, S., and Hallett, M. (2004). Identifying true brain interaction from EEG data using the imaginary part of coherency. *Clin. Neurophysiol.* 115, 2292–2307. doi: 10.1016/j.clinph.2004.04.029

Nunez, P. L., Srinivasan, R., Westdorp, A. F., Wijesinghe, R. S., Tucker, D. M., Silberstein, R. B., et al. (1997). EEG coherency: I: statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales. *Electroencephalogr. Clin. Neurophysiol.* 103, 499–515. doi: 10.1016/S0013-4694(97)00066-7

Schlögl, A., and Supp, G. (2006). Analyzing event-related EEG data with multivariate autoregressive parameters. *Prog. Brain Res.* 159, 135–147. doi: 10.1016/S0079-6123(06)59009-0

Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., and Leahy, R. M. (2011). Brainstorm: a user-friendly application for MEG/EEG analysis. *Comput. Intell. Neurosci.* 2011:879716. doi: 10.1155/2011/879716