Using Conditional Mutual Information to Approximate Causality for Multivariate Physiological Time Series

M.E. Pflieger, R.E. Greenblatt
Source Signal Imaging, Inc., San Diego, CA, USA

Abstract— **Causality analytic techniques based on conditional mutual information are described. Causality analysis may be used to infer linear and nonlinear causal relations between selected brain regions, and can account for identified non-causal confounds. The analysis results in a directed graph whose nodes are brain regions, and whose edges represent information flow. This causal information measure in principle should handle arbitrary nonlinear interactions without presupposing particular models of interaction**.

Keywords—**Functional connectivity, causality, multivariate time series analysis**

I. Introduction

There is a widespread belief that the brain is organized at the system level as a collection of semi-independent functional modules that dynamically establish interconnections during self-organization of cognition and behavior. This idea has emerged as an important paradigm for motivating research in systems, cognitive, and imaging neuroscience ([2], [4], [24]), and also for the understanding of clinical disorders such as schizophrenia ([1], [14], [35]) and autism ([16], [18]). Consequently, there is considerable interest in *effective connectivity analysis* [9] with the intent of teasing apart causal relationships in large-scale brain networks. Methods of effective connectivity analysis (typically applied to fMRI data) include structural equation modeling [23], nonlinear PCA [10], and coupled Volterra equations [4]. There are various methods for measuring *associations* between a pair of EMEG time series, and the most popular of these are linear measures, such as correlation. In the frequency domain, the phase of the linear coherence function may suggest a temporal lead or lag relationship between spontaneous EMEG or ECoG time series ([21], [33]). Analogous linear measures for event-related data are event-related covariance [11] and event-related coherence [32].

However, there are two salient limitations to measures like correlation and coherence. First, these measures of predictability are necessary but not sufficient for estimating causality, because they do not exclude possible non-causal confounds introduced, for example, by estimator crosstalk or volume conduction. Second, a limitation of all linear measures is the underlying assumption of Gaussian statistics, which is known to be violated in epilepsy [20], [26], and is probably invalid for many other areas of brain dynamics as well.

In this paper, we describe analytic techniques to infer linear and nonlinear causal relations between selected brain regions, after accounting for identified confounds. Causality analysis [12] can result in a directed graph whose nodes are brain regions, and whose edges represent information flow. This causal information measure in principle should handle arbitrary nonlinear interactions without presupposing particular models of interaction. Related measures applied to neurophysiological time series are described in [4], [6], [22], [26], [19], [8].

II. Methods

In order to estimate casual interaction between brain regions using data obtained non-invasively from EMEG, we must first establish a suitable method for estimating the state of selected brain regions of interest (ROIs), the state space representation. Then, causal interactions may be estimated using an extension of mutual information measures.

A. State space representation

Local linear estimators, such as REGAE [27] or LCMV beamformers [12] produce multivariate time series estimates for a brain region of interest (ROI) $\mathbf{r}(t) = [r_1(t) \cdots r_d(t)]^{\mathrm{T}}$, where d is the estimator dimension. If a differential equation governs the dynamics of the time series, then regardless of whether the differential equation is known, it is reasonable to think of the *dynamic state* of the time series at time t as the vector of time derivatives up to the order of the differential equation, $[r^{(0)}(t), \cdots, r^{(\kappa)}(t)]^{\mathrm{T}}$. Then the differential equation specifies the dynamic rule for updating the state of the time series. In practice, however, only the $r^{(0)}(t) = r(t)$ time series is available at discretely sampled intervals, with the consequence that it is required to estimate the higher order time derivatives from multiple time samples in a neighborhood of t. Moreover, real time series contain noise, and susceptibility to noise increases with the order of differentiation. Several methods may be applied successfully for state space representation that can overcome these problems, including delay vectors ([25], [31]), singular spectrum analysis [3], and complex demodulation [25].

B. Conditional mutual information (CMI)

Conditional mutual information (CMI) is the essential information theoretic concept used in defining causal information. CMI is based conceptually on probability densities on vector spaces, multivariate differential entropy, the Kullback-Leibler information divergence, and mutual information, each described in this section.

<u>*Probability densities on vector spaces*</u>. Let $p()$ be a given probability density function defined on a real or complex valued vector space, $\mathbf{X}$. If $\mathbf{X}_1$ and $\mathbf{X}_2$ are mutually orthogonal and complementary subspaces of $\mathbf{X}$ (i.e., $\mathbf{X}$ is the direct sum of $\mathbf{X}_1$ and $\mathbf{X}_2$, $\mathbf{X} = \mathbf{X}_1 \oplus \mathbf{X}_2$), then given $\mathbf{x} \in \mathbf{X}$ we may write $\mathbf{x}_1$ as the projection of $\mathbf{x}$ to $\mathbf{X}_1$,

$\mathbf{x}_2$ as the projection of $\mathbf{x}$ to $\mathbf{X}_2$, and $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)$. The marginal densities are

$$p(\mathbf{x}_1) \equiv \int_{\mathbf{X}_2} p(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_2, p(\mathbf{x}_2) \equiv \int_{\mathbf{X}_1} p(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_1 \quad (1)$$

$\mathbf{X}_1$ and $\mathbf{X}_2$ are p-independent if (and only if) $p()$ can be factored as $p(\mathbf{x}_1, \mathbf{x}_2) = p(\mathbf{x}_1) p(\mathbf{x}_2)$ for all $(\mathbf{x}_1, \mathbf{x}_2) \in \mathbf{X}$. The conditional density of $\mathbf{x}_1$ given $\mathbf{x}_2$ is

$$p(\mathbf{x}_1 \mid \mathbf{x}_2) \equiv \frac{p(\mathbf{x}_1, \mathbf{x}_2)}{p(\mathbf{x}_2)} \quad (2)$$

If $\mathbf{X}_1$ and $\mathbf{X}_2$ are p-independent, $p(\mathbf{x}_1 \mid \mathbf{x}_2) = p(\mathbf{x}_1)$.

Another form of equation (2) is $p(\mathbf{x}_1, \mathbf{x}_2) = p(\mathbf{x}_1 \mid \mathbf{x}_2) p(\mathbf{x}_2)$, which can be written more generally as the *chain rule* for obtaining the joint density as a product of marginal and conditional densities: If $\mathbf{X} = \mathbf{X}_1 \oplus \cdots \oplus \mathbf{X}_m$, then

$$\begin{aligned} &p(\mathbf{x}_1, \cdots, \mathbf{x}_m) = \\ &\quad p(\mathbf{x}_m) p(\mathbf{x}_{m-1} \mid \mathbf{x}_m) \cdots p(\mathbf{x}_1 \mid \mathbf{x}_2, \cdots, \mathbf{x}_m) \end{aligned} \quad (3)$$

In particular, if each component $\mathbf{x}_i = x_i$ is a scalar, note that each factor on the right-hand side is effectively a univariate density.

The *independence distribution* of $p()$ for a given vector space decomposition $\mathbf{X} = \mathbf{X}_1 \oplus \cdots \oplus \mathbf{X}_m$ is defined as $\dot{p}(\mathbf{x}_1, \cdots, \mathbf{x}_m) \equiv p(\mathbf{x}_1) p(\mathbf{x}_2) \cdots p(\mathbf{x}_m)$: $\dot{p}()$ selectively destroys all relationships that may exist in $p()$ between subspaces $\{\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_m\}$.

<u>*Multivariate differential entropy*</u>. The negative log density, $-\log p(\mathbf{x})$, measures the "unexpectedness" of observing $\mathbf{x}$ for a random draw from $p()$. (The particular base is not critical; we use the natural logarithm, base e). The unexpectedness of observing $\mathbf{x}$ approaches 0 as $p(\mathbf{x})$ approaches 1; it moves toward $+\infty$ as $p(\mathbf{x})$ approaches 0. The Shannon entropy ([7], [30]) is the expectation of the negative log density

$$H_p(\mathbf{X}) \equiv \mathrm{E}_p\left[-\log p(\mathbf{x})\right] = -\int_{\mathbf{X}} p(\mathbf{x}) \log p(\mathbf{x}) d\mathbf{x} \quad (4)$$

which measures "expected unexpectedness" or uncertainty.

Because the variables are continuous, this is known as *differential* entropy. By contrast with discrete entropy, differential entropy is not an absolute measure: It is relative to the coordinate system [30]. For example, the differential entropy can switch from positive to negative simply by changing measurement units: There is no absolute zero, and the entropy sign (positive or negative) has no special significance *eo ipso*. This "floating reference" issue may be addressed by explicitly introducing some known distribution $q()$ against which $p()$ is compared.

The <u>*Kullback-Leibler (KL) information divergence*</u> [19] of density $p()$ relative to density $q()$ is

$$D_{p\|q}(\mathbf{X}) \equiv \mathrm{E}_p\left[\log \frac{p(\mathbf{x})}{q(\mathbf{x})}\right] = -H_p(\mathbf{X}) - \mathrm{E}_p[\log q(\mathbf{x})] \quad (5)$$

which is a non-metric distance of $p()$ from $q()$: It is nonnegative, and zero only if $p() = q()$ (but is non-symmetric, and does not satisfy the triangle inequality). This measure is also known as relative entropy, noting that (5) differs from (4) by introducing $q()$ as a reference distribution (with sign reversal to produce nonnegative results). KL divergence measures the information available for discriminating the distribution of interest $p()$ from the reference distribution $q()$.

The <u>*conditional entropy*</u> of $\mathbf{X}_1$ on $\mathbf{X}_2$ is obtained by combining (2) and (4):

$$\begin{aligned} H_p(\mathbf{X}_1 \mid \mathbf{X}_2) &= \mathrm{E}_p\left[-\log \frac{p(\mathbf{x}_1, \mathbf{x}_2)}{p(\mathbf{x}_2)}\right] \\ &= H(\mathbf{X}_1, \mathbf{X}_2) - H(\mathbf{X}_2) \end{aligned} \quad . \quad (6)$$

Likewise, the <u>*conditional KL divergence*</u> of $\mathbf{X}_1$ on $\mathbf{X}_2$ combines (2) and (5):

$$\begin{aligned} D_{p\|q}(\mathbf{X}_1 \mid \mathbf{X}_2) &= \mathrm{E}_p\left[\log \frac{p(\mathbf{x}_1 \mid \mathbf{x}_2)}{q(\mathbf{x}_1 \mid \mathbf{x}_2)}\right] \\ &= D_{p\|q}(\mathbf{X}_1, \mathbf{X}_2) - D_{p\|q}(\mathbf{X}_2) \end{aligned} \quad (7)$$

<u>*Mutual information (MI)*</u>. In particular, given subspaces $\mathbf{X}_1$ and $\mathbf{X}_2$ (as above) we may choose $q()$ to be the *independence density* $q(\mathbf{x}_1, \mathbf{x}_2) = p(\mathbf{x}_1) p(\mathbf{x}_2)$. Then the mutual information of $\mathbf{X}_1$ and $\mathbf{X}_2$ in $p()$ is defined as the KL divergence of $p()$ relative to the independence density:

$$\begin{aligned} I_p(\mathbf{X}_1, \mathbf{X}_2) &\equiv D(p(\mathbf{x}_1, \mathbf{x}_2) \,\|\, p(\mathbf{x}_1) p(\mathbf{x}_2)) \\ &= \mathrm{E}_p\left[\log \frac{p(\mathbf{x}_1, \mathbf{x}_2)}{p(\mathbf{x}_1) p(\mathbf{x}_2)}\right] \\ &= H_p(\mathbf{X}_1) + H_p(\mathbf{X}_2) - H_p(\mathbf{X}_1, \mathbf{X}_2) \end{aligned} \quad (8)$$

[7]. Thus, mutual information increases with p-dependency and is zero only when $\mathbf{X}_1$ and $\mathbf{X}_2$ are p-independent.

<u>*Conditional Mutual Information (CMI)*</u>. We are now prepared to define the key concept of CMI. Let $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_3$ be mutually orthogonal and jointly complementary subspaces of $\mathbf{X}$ (i.e., $\mathbf{X} = \mathbf{X}_1 \oplus \mathbf{X}_2 \oplus \mathbf{X}_3$). Then the mutual information of $\mathbf{X}_1$ and $\mathbf{X}_2$ conditional on $\mathbf{X}_3$ in $p()$ is

$$\begin{aligned} &I_p(\mathbf{X}_1, \mathbf{X}_2 \mid \mathbf{X}_3) \equiv D(p(\mathbf{x}_1, \mathbf{x}_2 \mid \mathbf{x}_3) \,\|\, p(\mathbf{x}_1 \mid \mathbf{x}_3) p(\mathbf{x}_2 \mid \mathbf{x}_3)) \\ &= \mathrm{E}_p\left[\log \frac{p(\mathbf{x}_1, \mathbf{x}_2 \mid \mathbf{x}_3)}{p(\mathbf{x}_1 \mid \mathbf{x}_3) p(\mathbf{x}_2 \mid \mathbf{x}_3)}\right] \\ &= H_p(\mathbf{X}_1, \mathbf{X}_3) + H_p(\mathbf{X}_2, \mathbf{X}_3) - H_p(\mathbf{X}_3) - H_p(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3) \end{aligned}$$

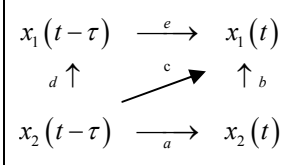
which reduces to ordinary mutual information (8) when $\mathbf{X}_3$ is p-independent of $\mathbf{X}_1$, $\mathbf{X}_2$, and $\mathbf{X}_1 \oplus \mathbf{X}_2$.

C. Causality

Let P_1 and P_2 be parallel processes with associated state spaces $\mathbf{X}_1$ and $\mathbf{X}_2$, respectively. A joint realization of the processes gives rise to two time series so that $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$ are the state vectors of the processes at time t. (For example, these may be the estimated states of activity in brain regions ROI 1 and ROI 2 at time t.) Based on observations of states of process P_1 and joint observations of states of process P_2 at relative time τ earlier (i.e., τ is a time lag from P_2 to P_1), we seek to quantify the information available at the earlier state of process P_2 about the later state of process P_1.

Predictive information (or *predictive mutual information*, PMI) is defined as the mutual information of states $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t-\tau)$, $I(\mathbf{X}_1(t), \mathbf{X}_2(t-\tau))$, which quantifies the predictability of the state of the first process having observed an earlier state of the second process. (Due to symmetry, it also quantifies "retrospective predictability" of the earlier second process state having observed a later first process state.)

Fig. 1. **Arrows *a* and *e* represent predictive self-information; arrows *b* and *d* represent zero-lag cross-information; and diagonal arrow *c* represents the predictive cross- information.**



Causal information (or *causal conditional mutual information*, CCMI) is defined as the conditional mutual information of states $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t-\tau)$ conditioned on states $\mathbf{x}_1(t-\tau)$ and $\mathbf{x}_2(t)$ — $I(\mathbf{X}_1(t), \mathbf{X}_2(t-\tau) \mid \mathbf{X}_1(t-\tau), \mathbf{X}_2(t))$ — which discounts as non-causal any predictability due to instantaneous between-process states and time-lagged within-process states (see Fig. 1). These may be estimated in a *linear* fashion by extracting the Gaussian information content, and in a *nonlinear* fashion by extracting residual information content after having removed the Gaussian content.

D. Numerical methods for entropy estimation

Linear Entropy. Let $\{\mathbf{x}_i \mid \mathbf{x}_i \in \mathbf{X}\}$ be a sample of N independent observations from an n-variate distribution with unknown density $p()$, and let $\hat{\mathbf{C}}$ be the $n \times n$ sample covariance matrix. Noting that entropy estimation is unaffected by the mean, we assume that the distribution has zero mean. Of all densities with covariance $\mathbf{C} = \hat{\mathbf{C}}$, the one with maximum entropy [30] is the n-variate Gaussian density

$$g(\mathbf{x}) \equiv (2\pi)^{-\frac{n}{2}} |\mathbf{C}|^{-\frac{1}{2}} \exp\{-\frac{1}{2}\mathbf{x}^*\mathbf{C}^{-1}\mathbf{x}\} \quad (9)$$

The entropy of a Gaussian distribution is

$$H_g(\mathbf{X}) = \frac{n}{2}\log(2\pi e) + \frac{1}{2}\sum_{i=1}^{n}\log[\mathbf{C}]_i \quad (10)$$

where the notation $[\mathbf{C}]_i$ here indicates the ith singular value (or eigenvalue) of $\mathbf{C}$. Thus, the Gaussian entropy may be estimated from the sample covariance matrix using purely linear computational methods, such as singular value decomposition: Gaussianity and linearity go hand-in-hand. The first term of (10) is a constant that does not depend on the particular sample $\{\mathbf{x}_i\}$; it depends only on the dimension n of space $\mathbf{X}$.

Recalling that differential entropy is relative to the coordinate system: Given different coordinate representations of the same sample, such as $\{\mathbf{x}_i\}$ and $\{\mathbf{x}'_j\}$, there is a linear transformation for each sample—the inverse square root of the sample covariance matrix—that converts all observations to a normalized, unitless system; namely $\hat{\mathbf{C}}^{-1/2}\mathbf{x}_j = \hat{\mathbf{C}}'^{-1/2}\mathbf{x}'_j$. This is descriptively called a *sphering* transformation because the sample covariance of the normalized observations is the identity matrix, $\mathbf{I}$, and thus the corresponding Gaussian distribution is hyper-spherical (about the origin) with standard deviation 1. Finally, we note that the sample-dependent second term of (10) is zero after conversion to normalized coordinates. Equivalently, the sphering transformation has equalized all variables while removing all linear relationships between them: After normalization, there is no further *linear information content* to be extracted from the sample. Thus, we define the linear information content of a sample as the second term of (10), which is the amount of entropy reduction after normalization.

Nonlinear Entropy. The problem that remains is to estimate the entropy of the sphered data sample, $\{\mathbf{z}_i \mid \mathbf{z}_i \in \mathbf{Z}\}$. One suitable numerical methods is the straightforward leave-one-out resubstitution method [15] that utilizes a Gaussian kernel probability density estimator:

$$H_n(\mathbf{Z}) = -\frac{1}{N}\sum_{i=1}^{N}\log p_i(\mathbf{z}_i)$$
$$= -\frac{1}{N}\sum_{i=1}^{N}\log\left\{\frac{1}{N-1}\sum_{j\neq i}\frac{1}{(2\pi)^{n/2}\sigma^n}\exp\left\{-\frac{|\mathbf{z}_j - \mathbf{z}_i|^2}{2\sigma^2}\right\}\right\}$$

The first summation estimates the expected value of $-\log p(\mathbf{z})$ based on the sample average of N observations. The probability density at each observed sample location is estimated as the average of Gaussian kernels (isotropic with standard deviation σ) centered on the *other* observed samples (excluding the location where the evaluation takes place). In a simulation study, we found empirically that $\sigma = 1$ is optimal for sphered data. If no exceptions are found—or, best case—if optimality of $\sigma = 1$ could be proved in general, then this estimator has no other free parameters: The nonlinear entropy (as well as the linear entropy) would be "unique" measures.

III. Conclusion

In conclusion, we have outlined a "successive approximation" approach to causality via conditional mutual information that discounts the effects of non-causal conditions. Preliminary results with these methods have been reported using simulated data [27] and also with clinical epilepsy data [28].

Acknowledgment

This work has been supported in part by NIH R43 EB000614.

References

[1] N.C. Andreasen *et al.*, "Schizophrenia and cognitive dysmetria: a positron-emission tomography study of dysfunctional prefrontal-thalamic-cerebellar circuitry," *PNAS*, vol. 19, pp. 9985-90, 1996.

[2] S.L. Bressler, "Large-scale cortical networks and cognition," *Brain Res Rev*, vol. 20, pp. 288-304, 1995.

[3] D.S. Broomhead, G.P. King, "Extracting quantitative dynamics form experimental data," *Physica D*, vol. 20, pp. 217-36, 1986.

[4] A. Brovelli *et al.*, "Beta oscillations in a large-scale sensorimotor network: Directional influences revealed by Granger causality," *PNAS*, vol. 101, pp. 9849-54, 2003.

[5] C. Buchel, K. Friston, "Interactions among neuronal systems assessed with functional neuroimaging," *Rev Neurol (Paris)*, vol. 157, pp. 807-15, 2001.

[6] S.A. Bunge, E. Hazeltine, M. Scanlon, A.C. Rosen, J.D.E. Gabrielli, "Dissociable contributions of prefrontal and parietal cortices to response selection," *NeuroImage,* vol. 17, pp. 1562-71, 2002.

[7] T.M. Cover, J.A. Thomas, *Elements of Information Theory*. New York: John Wiley, 1991.

[8] W.A. Freiwald *et al.*, "Testing non-linearity and directedness of interactions between neural groups in the macaque inferotemporal cortex," *J Neurosci Methods*, vol. 94, pp. 105-19, 1999.

[9] K. Friston, "Functional and effective connectivity: A synthesis," *Human Brain Mapping*, vol. 2, pp. 56-78, 1994.

[10] K. Friston, J. Phillips, D. Chawla, C. Buchel, "Revealing interactions among brain systems with nonlinear PCA," *Hum Brain Mapp*, vol. 8, pp. 92-7, 1999.

[11] A.S. Gevins *et al.*, Event-related covariances during a bimanual visuomotor task. I. Methods and analysis of stimulus- and response-locked data. *EEG clin Neurophysiol*. **74**:58-75, 1989.

[12] C.W.J. Granger, "Investigating causal relations by econometric methods and cross-spectral methods," *Econometrica*, vol. 37, pp. 424-38, 1969.

[13] R.E. Greenblatt, A. Ossadtchi, M.E. Pflieger, "Local linear estimators for the bioelectromagnetic inverse problem," *IEEE Trans Signal Proc*, in press.

[14] A.A. Ioannides, V. Poghosyan, J. Dummers, M. Streit, "Real-time neural activity and connectivity in healthy individuals and schizophrenia patients," *NeuroImage*, vol. 23, pp. 473-482, 2004.

[15] A.V. Ivanov, M.N. Rozhkova, "Properties of the statistical estimate of the entropy of a random vector with a probability density," *Problems of Information Transmission*, vol. 17, pp. 171-8, 1981.

[16] M.A. Just, V.L. Cherkassky, T.A. Keller, N.J. Minshew, "Cortical activation and synchronization during sentence comprehension in high-functioning autism: evidence of underconnectivity," *Brain*, vol. 127, pp. 1811-21, 2004.

[17] J.A.S. Kelso, *Dynamical Patterns: The Self-Organization of Brain and Behavior*. Cambridge, MA: MIT Press, 1995.

[18] H. Koshino *et al*., "Functional connectivity in an fMRI working memory task in high-functioning autism," *NeuroImage*, vol. 24, pp. 810-21, 2005.

[19] S. Kullback, *Information Theory and Statistics*. Mineola, NY: Dover, 1959 (reprinted 1997).

[20] M. Le Van Quyen, C. Adam, M. Baulac, J. Martinerie, F.J. Varela, "Nonlinear interdependencies of EEG signals in human intracranially recorded temporal lobe seizures," *Brain Res*, vol. 792, pp. 24-40, 1998.

[21] J.P. Lieb, K. Hoque, C.E. Skomer, X.W. Song XW, "Inter-hemispheric propagation of human mesial temporal lobe seizures: a coherence/phase analysis," *EEG clin Neurophysiol*, vol. 67, pp. 101-19, 1987.

[22] N.J.I. Mars, F.H. Lopes da Silva FH, "EEG analysis methods based on information theory," in *Handbook of Electroencephalography and Clinical Neurophysiology, Revised Series, Vol. 1*, A.Ss Gevins, A. Rémond, Eds. Amsterdam: Elsevier, 1987, pp. 297-307.

[23] A.R. McIntosh, F. Gonzalez-Lima, "Structural equation modeling and its application to network analysis in functional brain imaging," *Hum Brain Mapp*, vol. 2, 2-22, 1994.

[24] M.M. Mesulam, "Large-scale neurocognitive networks and distributed processing for attention, language, and memory," *Ann Neurol*, vol. 28, pp. 597-613, 1990.

[25] R.K. Otnes, L. Enochson, *Applied Time Series Analysis*. New York: John Wiley, 1978.

[26] J.P.M. Pijn JPM, *Quantitative Evaluation of EEG Signals in Epilepsy: Nonlinear Associations, Time Delays, and Nonlinear Dynamics*. Amsterdam: Rodopi, 1990.

[27] M.E. Pflieger ME, "Time-lagged causal information: a new metric for effective connectivity analysis," *NeuroImage*, vol. 19, CD-ROM, 2003.

[28] M.E. Pflieger, B.A. Assaf, "A non-invasive method for analysis of epileptogenic brain activity," *Epilepsia*, vol. 45, suppl 7, pp. 70-1, 2004.

[29] M.E. Pflieger, R.E. Greenblatt, J. Kirkish, "Regional resolving power of combined MEG/EEG," *Neurol Clin Neurophysiol*, in press.

[30] C.E. Shannon, W. Weaver, *The Mathematical Theory of Communication*. Urbana, IL: University of Illinois Press, 1949.

[31] F. Takens, "Detecting strange attractors in turbulence," in *Proc. Symp. Dynamical Systems and Turbulence*, D.A. Rand, L.S. Young, Eds. Berlin: Springer, 1981.

[32] R.W. Thatcher, C. Toro, M.E. Pflieger, M. Hallett, "Human neural network dynamics using multimodal registration of EEG, PET, and MRI," *Functional Neuroimaging, Technical Foundations*, R.W. Thatcher *et al.*, Eds. San Diego: Academic Press, pp. 269-278, 1994.

[33] V.L. Towle, R.K. Carder, L. Khorasani, D. Lindberg, "Electrocorticographic coherence patterns," *J Clin Neurophysiol*, vol. 16, pp. 528-47, 1999.

[34] F. Varela, J-P. Lachaux, E. Rodriguez, J. Martinerie, "The Brainweb: phase synchronization and large-scale integration," *Nature Rev Neuroscience*, vol. 2, pp. 229-39, 2001.

[35] G. Winterer, R. Coppola, M.F. Egan, T.E. Goldberg, D.R. Weinberger, "Functional and effective frontotemporal connectivity and genetic risk for schizophrenia," *Biol Psychiatry*, vol. 54, pp. 1181-92, 2003.