VARX model of the Brain: Functional connectivity dominates response to natural stimulus in human intracranial EEG

Maximilian Nentwich [1], **Behtash Babadi** [2]
Stephan Bickel [1] **Lucas C. Parra** [3]
1 The Feinstein Institutes for Medical Research, Northwell Health
2 Electrical and Computer Engineering, University of Maryland
3 Biomedical Engineering, City College of New York

Abstract

When ongoing sensory stimulation reaches the brain, the activity that it generates reverberates across recurrent brain networks. To distinguish the effect of the stimulus from the recurrent brain dynamic we propose a vector-autoregressive model with external input (VARX). This model decomposes the traditional multivariate "temporal response functions" into an immediate response to the stimulus, and a recurrent dynamic within the brain. The model naturally combines with the statistical formalism of Granger to determine statistical significance of each of these effects, including direction of effects between brain areas based on temporal precedence. We demonstrate the approach with intracranial EEG (iEEG) recordings of human subjects watching movies (22 patients). We find that the "recurrent connectivity" during the movies is nearly identical when subjects are at rest. The recurrent dynamic enhances and prolongs the responses to sound, scene cuts and other external events. If these external inputs are not accounted for, they induce spurious connectivity between brain areas. The recurrent connectivity differs from conventional "functional connectivity" in that it is directed and asymmetric. We find that sensory areas have mostly outward connections, whereas higher-order areas have mostly incoming connections. The proposed VARX model combines notions of functional connectivity with stimulus response, potentially resolving conceptual problems that arise when discussing temporal integration or prediction whilst ignoring the recurrent dynamic of the brain. Code for the proposed VARX Granger analysis is available here `https://github.com/lcparra/varx`.

Introduction

The brain is highly interconnected between and within brain areas. Strikingly, most models of brain activity in response to an external natural stimuli do not take the recurrent architecture of brain networks into account. "Encoding" models often rely on simple input/output relationships such as general linear models in fMRI (Friston et al., 1995), or temporal response functions in EEG/MEG (Lalor & Foxe, 2010). Interactions between brain areas are commonly captured during rest, often simply as instantaneous linear correlations, referred to as "functional connectivity". (Greicius, Krasnow, Reiss, & Menon, 2003). By taking temporal precedence into account with linear predictive models the "Granger-causality" formalism establishes directed "connectivity" (Haufe, Nikulin, Müller, & Nolte, 2013; Sheikhattar et al., 2018; Soleimani et al., 2022; Pellegrini, Delorme, Nikulin, & Haufe, 2023).

These concepts all naturally combine into a vector-autoregressive model with external input (VARX), which is well established in the field of linear systems (Ljung, 1999) and econometrics (Hamilton, 2020). While linear systems are an inadequate model of neuronal dynamics, they remain an important tool to understand neural representations because of their conceptual simplicity. They are routinely used to link non-linear features of continuous stimuli to neural responses such as video to fMRI (Naselaris, Kay, Nishimoto, & Gallant, 2011) or speech to EEG (Di Liberto, O'Sullivan, & Lalor, 2015). They are used even to understand responses of deep-neural network models (Li et al., 2023). Here we use this classic linear formalism to combine two canonical concepts which have thus far largely remained separated, namely, that of stimulus encoding and that of the internal brain dynamics.

Methods

The VARX model explains a time-varying vectorial signal $\mathbf{y}(t)$ as the result of an autoregressive feedback driven by an innovation process $\mathbf{e}(t)$ and an external input $\mathbf{x}(t)$ (Ljung, 1999):

$$\mathbf{y}(t) = \mathbf{A} * \mathbf{y}(t-1) + \mathbf{B} * \mathbf{x}(t) + \mathbf{e}(t) \tag{1}$$

$\mathbf{A}*$ and $\mathbf{B}*$ represent convolutions with appropriately sized filter matrices. They can be identified by minimizing the mean squared innovation, $\sigma^2 = 1/T \sum_{t=1}^{T} \mathbf{e}(t)^2$. This is also the prediction error, for predicting $\mathbf{y}(t)$ from the history $\mathbf{y}(t-1)$ and input $\mathbf{x}(t)$. In the Granger formalism this error is calculated with all predictors included (full model, σ_f) or with individual channels in $\mathbf{y}(t-1)$ or $\mathbf{x}(t)$ omitted (reduced model, σ_r) (Granger, 1969). To quantify the "effect" of the specific channel one can take the ratio of these errors (Geweke, 1982) leading to the test statistic D known as the "deviance". For large T deviance follows the Chi-square distribution with cumulative density F, from which one can compute the p-value:

$$D = T \log(\sigma_r^2/\sigma_f^2) \tag{2}$$

$$p = 1 - F(D,T) \tag{3}$$

$$R^2 = 1 - \exp(-D/T) \tag{4}$$

The "generalized" R^2 (Magee, 1990) serves as a measure of effect size. A full description of the approach can be found in (Parra, Silvan, Nentwich, Madsen, & Babadi, 2024).

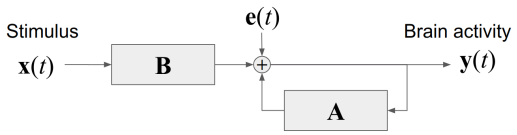

Figure 1: **VARX model:** $\mathbf{A}$ captures recurrent connectivity between brain areas and $\mathbf{B}$ captures response to external inputs. The overall brain response to the stimulus is given by the system impulse response $\mathbf{H} = \mathbf{B}/(1-\mathbf{A})$. $\mathbf{e}(t)$ captures the innovation in brain activity, and is assumed to be uncorrelated.

Results

We used this VARX Granger analysis on the intracranial EEG data from humans (N=22) (Nentwich et al., 2023). We first compared the effect size R for the recurrent connectivity $\mathbf{A}$ during movie watching (Fig. 2A) and eyes-open rest (Fig. 2B). Surprisingly, the median difference in R is unchanged (Fig. 2C&D, student t-test, t(21)=-1.38, p=0.18).

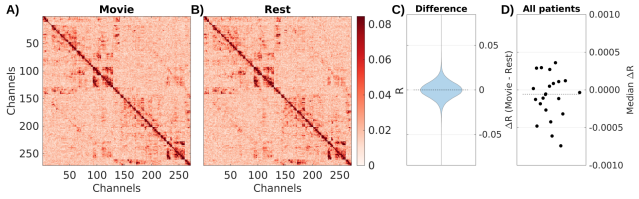

Figure 2: **Recurrent connectivity $\mathbf{A}$ during movies does not differ from rest.** A) VARX model fit on 5 minutes of raw iEEG data, with sound envelope, fixation onsets and film cuts as input features. B) VARX model during resting fixation with fixation onset as input feature. A) and B) show the coefficient of determination R for each connection in $\mathbf{A}$. C) Difference of R between movies and rest for one subject; D) for all subjects.

If we ignore the stimulus features during movie watching, by only fitting a VAR model, there is a significantly higher ratio of recurrent connections (Fig. 3C, t(21)=24.53, p<.0001). In other words, there are spurious recurrent connections due to stimulus-induced correlation. However, the effect size significantly decreases (Fig. 3D, t(21)=-9.48, p<.0001). In short, accounting for the external input results in a model with stronger and more sparse recurrent connections.

We also compared the total system response $\mathbf{H}$, which includes the effect of recurrent connections $\mathbf{A}$, with the immediate effect $\mathbf{B}$, which does not (Fig. 4). We see that the total response is much stronger and longer (Fig. 4D, t(21)=-9.57, p< 0.0001). Incidentally, we find that our estimate of $\mathbf{H}$ is nearly identical to the estimate of the "multivariate temporal response function" (Crosse, Di Liberto, Bednar, & Lalor, 2016) (mTRF, not shown). In other words, $\mathbf{B}$ and $\mathbf{A}$ are a valid decomposition of the mTRF into immediate and recurrent effects.

Finally, we note the recurrent connections are potentially asymmetric (Fig. 5A), indicating that preceding activity in one

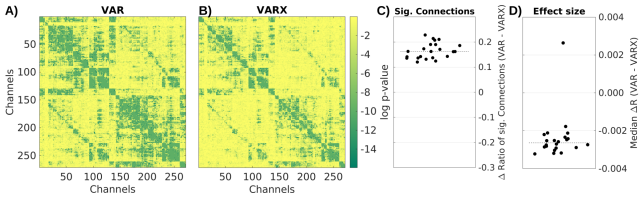

Figure 3: **Spurious recurrent connectivity in $\mathbf{A}$ is accounted for when modeling effect of input with $\mathbf{B}$.** A) p-values for each connection in $\mathbf{A}$ for VARX model on one subject; B) for VAR model; Both models are fit to the same data. C) Difference of fraction of significant recurrent connections between VAR and VARX for all subjects D) Median difference in R between VAR and VARX over all electrodes and subjects.

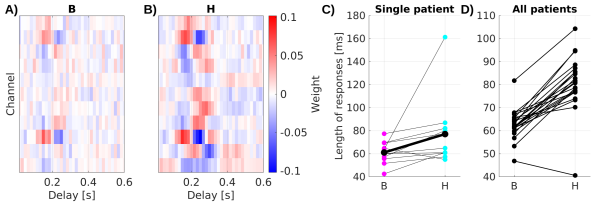

Figure 4: **Impulse response models.** A) Immediate responses $\mathbf{B}$ are weaker and shorter than the overall system response $\mathbf{H}$. B). C) Response length (full-width-half-max of the peak). D) Mean length of responses (22 patients).

brain area is a better predictor of activity in another, compared to the reverse direction, i.e. the two directions of the "Granger-causal" effect are not equally strong. We measured this asymmetry for each channel (N=4408). Grouping channels by cortical parcellations we find that effects in sensory areas (transverse temporal and fusiform gyrus) are mostly outgoing toward other brain areas, while in higher-order areas (precuneus, anterior cingulate, orbitofrontal cortex) effects are mostly incoming (Fig. 5B).

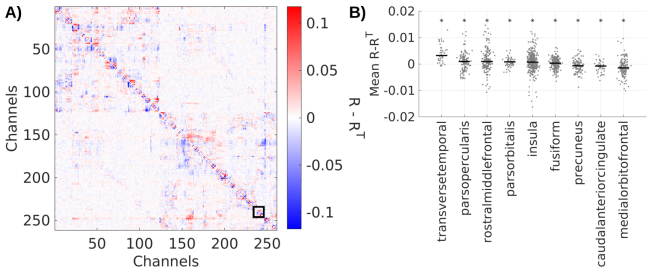

Figure 5: **Recurrent connectivity is directed from sensory to higher order areas.** A) Difference of $R - R^\top$ showing asymmetric directed effects. B) Directionality for each channel (mean along columns in $R - R^\top$) is significantly different from zero in 9 out of 36 brain areas in the Desikan-Killiany atlas.

Acknowledgments

We thank Charlie Schroeder for useful discussions. The work was supported by NIH grant P50 MH109429.

References

Crosse, M. J., Di Liberto, G. M., Bednar, A., & Lalor, E. C. (2016, November). The multivariate temporal response function (mTRF) toolbox: A MATLAB toolbox for relating neural signals to continuous stimuli. *Front. Hum. Neurosci.*, *10*, 604.

Di Liberto, G. M., O'Sullivan, J. A., & Lalor, E. C. (2015, October). Low-frequency cortical entrainment to speech reflects phoneme-level processing. *Curr. Biol.*, *25*(19), 2457–2465.

Friston, K. J., Holmes, A. P., Poline, J. B., Grasby, P. J., Williams, S. C., Frackowiak, R. S., & Turner, R. (1995, March). Analysis of fMRI time-series revisited. *Neuroimage*, *2*(1), 45–53.

Geweke, J. (1982, June). Measurement of linear dependence and feedback between multiple time series. *J. Am. Stat. Assoc.*, *77*(378), 304–313.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, 424–438.

Greicius, M. D., Krasnow, B., Reiss, A. L., & Menon, V. (2003, January). Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proc. Natl. Acad. Sci. U. S. A.*, *100*(1), 253–258.

Hamilton, J. D. (2020). *Time series analysis*. Princeton university press.

Haufe, S., Nikulin, V. V., Müller, K.-R., & Nolte, G. (2013, January). A critical assessment of connectivity measures for EEG data: a simulation study. *Neuroimage*, *64*, 120–133.

Lalor, E. C., & Foxe, J. J. (2010, January). Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution. *Eur. J. Neurosci.*, *31*(1), 189–193.

Li, Y., Anumanchipalli, G. K., Mohamed, A., Chen, P., Carney, L. H., Lu, J., ... Chang, E. F. (2023, October). Dissecting neural computations in the human auditory pathway using deep neural networks for speech. *Nat. Neurosci.*.

Ljung, L. (1999). System identification-theory for the user 2nd edition ptr prentice-hall. *Upper Saddle River, NJ*.

Magee, L. (1990, August). R2Measures based on wald and likelihood ratio joint significance tests. *Am. Stat.*, *44*(3), 250–253.

Naselaris, T., Kay, K. N., Nishimoto, S., & Gallant, J. L. (2011, May). Encoding and decoding in fMRI. *Neuroimage*, *56*(2), 400–410.

Nentwich, M., Leszczynski, M., Russ, B. E., Hirsch, L., Markowitz, N., Sapru, K., ... Parra, L. C. (2023, May). Semantic novelty modulates neural responses to visual change across the human brain. *Nat. Commun.*, *14*(1), 2910.

Parra, L. C., Silvan, A., Nentwich, M., Madsen, J., & Babadi, B. (2024). *Varx granger analysis: Modeling, inference, and applications.*

Pellegrini, F., Delorme, A., Nikulin, V., & Haufe, S. (2023). Identifying good practices for detecting inter-regional linear functional connectivity from eeg. *NeuroImage*, *277*, 120218.

Sheikhattar, A., Miran, S., Liu, J., Fritz, J. B., Shamma, S. A., Kanold, P. O., & Babadi, B. (2018, April). Extracting neuronal functional network dynamics via adaptive granger causality analysis. *Proc. Natl. Acad. Sci. U. S. A.*, *115*(17), E3869–E3878.

Soleimani, B., Das, P., Dushyanthi Karunathilake, I. M., Kuchinsky, S. E., Simon, J. Z., & Babadi, B. (2022, October). NLGC: Network localized granger causality with application to MEG directional functional connectivity analysis. *Neuroimage*, *260*(119496), 119496.