Intrinsic dynamic shapes responses to external stimulation in the human brain

Maximilian Nentwich[1], Marcin Leszczynski[2,3,4], Charles E. Schroeder[2,3], Stephan Bickel[1,5,6], Lucas C. Parra[7]

[1] The Feinstein Institutes for Medical Research, Northwell Health, Manhasset, NY, USA; [2] Departments of Psychiatry and Neurology, Columbia University College of Physicians and Surgeons, New York, NY, USA; [3] Translational Neuroscience Lab Division, Center for Biomedical Imaging and Neuromodulation, Nathan Kline Institute, Orangeburg, NY, USA; [4] Cognitive Science Department, Institute of Philosophy, Jagiellonian University, Kraków, Poland; [5] Departments of Neurology and Neurosurgery, Zucker School of Medicine at Hofstra/Northwell, Hempstead, NY, USA; [6] Center for Biomedical Imaging and Neuromodulation, Nathan Kline Institute, Orangeburg, NY, USA; [7] Department of Biomedical Engineering, The City College of New York, New York, NY, USA

Summary

Sensory stimulation of the brain reverberates in its recurrent neural networks. However, current computational models of brain activity do not separate immediate sensory responses from this intrinsic dynamic. We apply a vector-autoregressive model with external input (VARX), combining the concepts of “functional connectivity” and “encoding models”, to intracranial recordings in humans. This model captures the extrinsic effect of the stimulus and separates that from the intrinsic effect of the recurrent brain dynamic. We find that the intrinsic dynamic enhances and prolongs the neural responses to scene cuts, eye movements, and sounds. Failing to account for these extrinsic inputs, leads to spurious recurrent connections that govern the intrinsic dynamic. We also find that the recurrent connectivity during rest is reduced during movie watching. The model shows that an external stimulus can reduce intrinsic noise. It also shows that sensory areas have mostly outward, whereas higher-order brain areas mostly incoming connections. We conclude that the response to an external audiovisual stimulus can largely be attributed to the intrinsic dynamic of the brain, already observed during rest.

Keywords:
Connectivity, Granger analysis, VAR, recurrent networks, encoding models, intracranial EEG, eye movements, naturalistic

Author contributions:
Conceptualization, L.C.P. and M.N.; Methodology, L.C.P.; Software, L.C.P. and M.N.; Formal Analysis: M.N and L.C.P.; Investigation, M.L. and M.N.; Resources, S.B.; Writing – Original Draft, L.C.P. and M.N.; Writing – Review & Editing, L.C.P., M.N., M.L., C.E.S., S.B.; Funding Acquisition, L.C.P., C.E.S. and S.B.; Visualization, M.N. and L.C.P.; Supervision, L.C.P., C.E.S., S.B.

Declaration of interests:
The authors declare no competing interests.

Supplemental information:
Figures S1–S14 and Table S1

Data and code availability:
The raw data reported in this study cannot be deposited in a public repository because of patient privacy concerns. To request access, contact The Feinstein Institutes for Medical Research, through Dr. Stephan Bickel. In addition, processed datasets derived from these data have been deposited at https://doi.org/10.17605/OSF.IO/VC25T.

All original code has been deposited at https://github.com/MaxNentwich/varx_demo and is publicly available at https://doi.org/10.5281/zenodo.15127333.

Introduction

The primate brain is highly interconnected between and within brain areas. This includes areas involved in sensory processing [1]. Strikingly, most computational models of brain activity in response to external natural stimuli do not take the recurrent architecture of brain networks into account. We will refer to the dynamic driven by this recurrent architecture as the *intrinsic dynamic* of the brain. "Encoding" models often rely on simple input/output relationships such as general linear models in fMRI [2], or temporal response functions in EEG/MEG [3]. Interactions between brain areas are captured often just as instantaneous linear correlations that are referred to as "functional connectivity" when analyzing fMRI activity [4]. Others capture synchronous activity in different brain areas by measuring phase locking of electrical neural signals [5]. However, these measures of instantaneous correlation do not capture time delays inherent in recurrent connections. By taking temporal precedence into account with recurrent models the "Granger-causality" formalism can establish directed "connectivity". This has been used to analyze both fMRI and electrical activity [6–11].

The concept of functional connectivity was first developed to analyze neural activity during rest, where there are no obvious external signals to stimulate brain activity. But it is now also used to analyze brain activity during passive exposure to a stimulus, such as watching movies [12–15]. A general observation of these studies is that a portion of the functional connectivity is preserved between rest and stimulus conditions, while some aspects are altered by the perceptual task [12,16], sometimes showing increased connectivity during the stimulus.[15] This should be no surprise, given that an external stimulus can drive multiple brain areas and thus induce correlations between these areas [17]. Removing such stimulus-induced correlations by controlling for a common cause is standard practice in statistical modeling and causal inference [18]. However, in studies that focus on functional connectivity in neuroscience, stimulus-induced correlations are often ignored when analyzing the correlation structure of neural signals. A notable exception is “dynamic causal modeling” [19]. In this modeling approach the “input” can modulate functional connectivity. This is particularly important in the context of active behavioral tasks, where the common finding is that correlation structure changes with task states [20].

In this study we are interested in “passive” tasks such as rest and movie watching. We will ask here whether, after removing some of the stimulus-induced correlations, the intrinsic dynamic is similar between stimulus and rest conditions. Attempts to factor out the effects of intrinsic dynamics from that of the stimulus come from work on response variability. For instance, fMRI shows that variability across trials in motor cortex is due to an intrinsic “noise” which is linearly superimposed on a common response in both hemispheres due to a simple motor action [21]. Stimulus-response variability in the visual cortex has been attributed to variability of the ongoing dynamic [22,23]. Some studies of electrical recordings from the visual cortex show that correlations of spiking activity between different recording locations are largely unaffected by visual stimulation [24]. Yet, other studies show that visual input affects local correlation in the visual cortex [25–27] and across the brain [28].

The technical challenge when addressing these questions is to separate the direct effect of the stimulus from the intrinsic dynamic. Here we propose to separate these effects by modeling them simultaneously with the simplest possible model, namely, linear intrinsic effects between brain areas and linear responses to extrinsic input. A mathematical model that implements this is the vector-autoregressive model with external input (VARX). This model is well established in the field of linear systems [29] and econometrics [30], where it is used to capture intrinsic dynamics in the presence of an external input. The VARX model is an extension of the VAR model that is routinely used to establish "Granger-causality" in neuroscience (cited above). In the VARX model, Granger analysis provides a measure of statistical significance for the external input, as well as for the intrinsic dynamic, including its directionality, all as part of a single model.[31]

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 for event-related fMRI analysis but also for “encoding models” to link non-linear features of continuous stimuli to neural responses. They have been used to analyze responses to video in fMRI [32], to speech in EEG [33] or to audio in intracranial EEG [34]. They are even used to analyze the encoding in deep-neural network models [35]. Here we use a classic linear model to combine two canonical concepts in neuroscience, which have thus far remained separated, namely, that of “encoding models” [32] and “functional connectivity” models [6]. We will use this to analyze whole-brain, intracranial EEG in human patients at rest, and while they watch videos. Our main finding is that the recurrent connections observed during rest are only minimally altered by watching videos. Instead, the brain's response to naturalistic stimulus appears to be substantially shaped by the same intrinsic dynamic of the brain observed during rest.

Methods

The vector-autoregressive model with external input (VARX) falls within a group of well-established linear models used in neuroscience (see Table 1). Prominent examples in this group are the generalized linear model (GLM), dynamic causal model (DCM) and temporal response functions (TRF). While these models have been extensively used for neural signal analysis, the VARX model has not. We start therefore with a brief introduction. For more details please refer to [31].

VARX model

The VARX model explains a time-varying vectorial signal $\mathbf{y}(t)$ as the result of an intrinsic autoregressive feedback driven by an innovation process $\mathbf{e}(t)$ and an extrinsic[1] input $\mathbf{x}(t)$. For the ith signal channel the recurrence of the VAX model is given by:

[1] We adopt here the terminology of “intrinsic” and “extrinsic” as it is commonly used in neuroscience and psychology. In system modeling and econometrics, where the VARX model is prevalent, the more common terminology is “endogenous” and “exogenous”, meaning the same things.

$$\mathbf{y}_i(t) = \sum_{j=1}^{d_y} \sum_{\tau=1}^{n_a} \mathbf{A}_{ij}(\tau)\mathbf{y}_j(t-\tau) + \sum_{j=1}^{d_x} \sum_{\tau=0}^{n_b} \mathbf{B}_{ij}(\tau)\mathbf{x}_j(t-\tau) + \mathbf{e}_i(t)$$

$\mathbf{A}$ and $\mathbf{B}$ are matrices of filters of lengths n_a and n_b respectively. Therefore, $\mathbf{A}$ has dimensions $[d_y, d_y, n_a]$ and $\mathbf{B}$ has dimensions $[d_y, d_x, n_b]$, where d_y, d_x are the dimensions of $\mathbf{y}(t)$ and $\mathbf{x}(t)$ respectively. The innovation process $\mathbf{e}(t)$ captures the internal variability of the model. Without it, repeating the same input $\mathbf{x}(t)$ would always result in a fixed deterministic output $\mathbf{y}(t)$. The innovation is assumed to be uncorrelated in time and has therefore an uniform spectrum. The recurrent filters in $\mathbf{A}$ modify this spectrum to match the spectrum of $\mathbf{y}(t)$, thereby capturing the intrinsic dynamic. The feed-forward filters in $\mathbf{B}$ inject a filtered version of the extrinsic input $\mathbf{x}(t)$ into this intrinsic dynamic. The role of each of these terms for brain activity is explained in Fig. 1. We will refer to the filters in matrix $\mathbf{A}$ and $\mathbf{B}$ as recurrent and feed-forward "connections", but avoid the use of the word "causal" which can be misleading.[31]

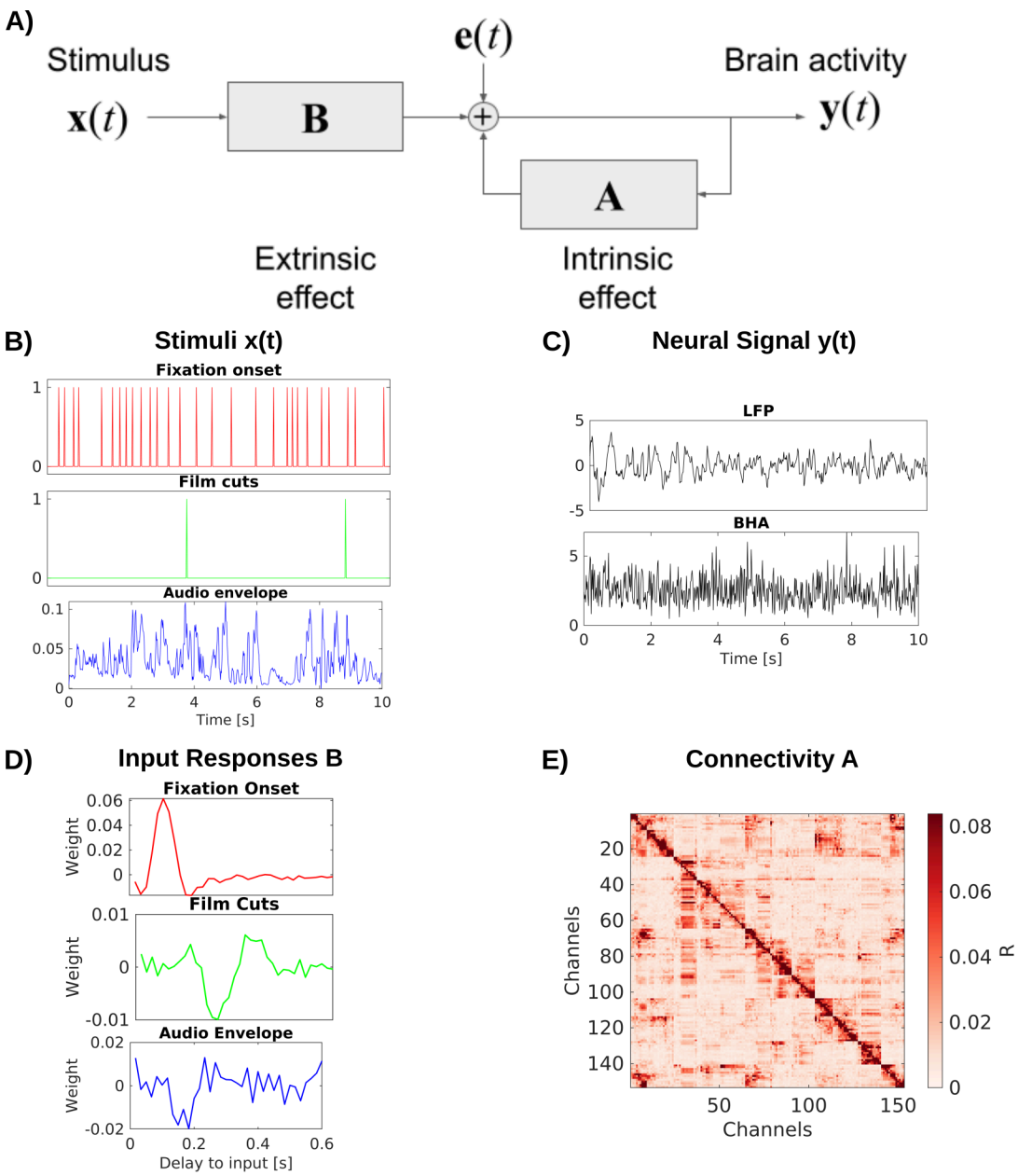

Figure 1: VARX model of the brain: A) Block diagram of the VARX model. $\mathbf{y}(t)$ represents observable neural activity in different brain areas, $\mathbf{x}(t)$ are observable features of a continuous sensory stimulus, $\mathbf{A}$ represent the recurrent connections within and between brain areas (intrinsic effect), and $\mathbf{B}$ captures the transduction of the sensory stimuli into neural activity and transmission to different brain areas (extrinsic effect). The diagonal term in $\mathbf{A}$ captures recurrent feedback within a brain area. Finally, $\mathbf{e}(t)$ captures unobserved "random" brain activity, which leads to intrinsic variability. B) Example of input stimulus features x(t). C) Example of neural signal y(t) recorded at a single location in the brain. We analyze local field potentials (LFP) and broad-band high frequency activity (BHA) in separate analyses. D) Examples of filters $\mathbf{B}$ for individual feed-forward connections between an extrinsic input and a specific recording location in the brain. E) Effect size R for the recurrent connections captured by auto-regressive filters $\mathbf{A}$.

Filter matrices $\mathbf{A}$ and $\mathbf{B}$ are unknown and can be estimated from the observed history of $\mathbf{x}(t)$ and $\mathbf{y}(t)$ using ordinary least squares (OLS). The objective for the optimal model is to minimize the power of the unobserved innovation process $\mathbf{e}(t)$, i.e. the summed squares:

$$\boldsymbol{\sigma}^2 = 1/T \sum_{t=1}^{T} \mathbf{e}(t)^2$$

Granger analysis

The innovation is also the prediction error, for predicting $\mathbf{y}(t)$ from the past $\mathbf{y}(t-1)$ and input $\mathbf{x}(t)$. In the Granger formalism the prediction error is calculated with all predictors included (error of the full model, $\boldsymbol{\sigma}_f$) or with individual dimension in $\mathbf{y}(t-1)$ or $\mathbf{x}(t)$ omitted from the prediction (error of the reduced models, $\boldsymbol{\sigma}_r$) [36]. To quantify the "effect" of the specific dimension one can take the ratio of these errors [37] leading to the test statistic D known as the "deviance". When the number of samples T is large, the deviance follows the Chi-square distribution with cumulative density F, from which one can compute a p-value:

$$\begin{aligned} D &= T \log(\boldsymbol{\sigma}_r^2 / \boldsymbol{\sigma}_f^2) \\ p &= 1 - F(D, T) \\ R^2 &= 1 - \exp(-D/T) \end{aligned}$$

The p-value quantifies the probability that a specific connection in either $\mathbf{A}$ or $\mathbf{B}$ is zero. Therefore, D, p and R^2 all have dimensions $[d_y, d_y]$ or $[d_y, d_x]$ for $\mathbf{A}$ or $\mathbf{B}$ respectively. respectively. The "generalized" R^2 [38] serves as a measure of effect size, capturing the strength of each connection. While this Granger formalism is well established in the context of estimating $\mathbf{A}$, i.e. VAR models, to our knowledge, it has not been used in the context of estimating $\mathbf{B}$, i.e. VARX or TRF models.

Overall system response

The overall brain response to the stimulus for the VARX model is given by the system impulse response (written here in the z-domain, or Fourier domain):

$$\mathbf{H} = (1 - \mathbf{A})^{-1}\mathbf{B}.$$

What we see here is that the system response $\mathbf{H}$ is factorized into an autoregressive (AR) filter $\mathbf{A}$ and a moving average (MA) filter $\mathbf{B}$. When modeled as a single MA filter, the total system response has been called the “multivariate Temporal Response Function” (mTRF) in the neuroscience community [39]. We found that the VARX estimate of $\mathbf{H}$ is nearly identical to the estimated mTRF [31]. In other words, $\mathbf{B}$ and $\mathbf{A}$ are a valid factorization of the mTRF into feed-forward extrinsic versus recurrent intrinsic effects.

Note that the extrinsic effects captured with filters $\mathbf{B}$ are specific (every stimulus dimension has a specific effect on each brain area), whereas the intrinsic dynamic propagates this initial effect to all connected brain areas via matrix $\mathbf{A}$, effectively mixing and adding the responses of all stimulus dimensions. Therefore, this factorization separates stimulus-specific effects from the shared intrinsic dynamic.

Relation to common neural signal models

The VARX model fits naturally into the existing family of models used for neural signals analysis (Table 1). While they differ in the formulation and statistical assumptions, their defining equations have a similar general form with the following attributes:

Table 1: Models commonly used in neural signal analysis

Model	Intrinsic effect **A**	Extrinsic effect **B**	Interact	Delay n_a, n_b	Estimation speed	Reference, with code where available
GLM	no	yes	no	=1	medium	[40], SPM, FSL
DCM	yes	yes	yes[2]	=1[3]	slow	[19], no code
VAR	yes	no	no	>1	fast/slow	[41]
mTRF	no	yes	no	>1	fast	[39]
VARX	yes	yes	no[4]	>1	fast	[31]

An important simplifying assumption for the mTRF, VAR, and VARX models is that $\boldsymbol{y}(t)$ is observable with additive normal distributed innovation. As a result, parameter estimation can use ordinary least squares, which is fast to compute. In contrast, GLM, DCM, and some variants of VAR models assume that $\boldsymbol{y}(t)$ is not directly observable, and needs to be estimated in addition to the unknown parameters $\mathbf{A}$ or $\mathbf{B}$. The same is true for the basic "output error" model in linear systems theory [29]. This requires slower iterative algorithms, such as expectation maximization. As a result, these models are often limited to small networks[5] of a few nodes to test specific alternative hypotheses [42]. In contrast, here we will analyze up to 300 channels per patient to draw general conclusions about overall brain organization.

Validation of recurrent connectivity estimate with whole-brain neural mass model

To test the descriptive validity [43] of the VARX model we follow the approach of recovering structural connectivity from functional activity in simulation [44]. Specifically, we will compare the recurrent connectivity $\mathbf{A}$ derived from brain activity simulated assuming a given structural

---

[2] "Interact" refers to an additional bilinear interaction term of the form $\boldsymbol{x}\,\boldsymbol{C}\,\boldsymbol{y}$ that allows for a modulation of intrinsic effect by the external input.
[3] The DCM is defined in terms of the first derivative of $\boldsymbol{y}(t)$, which in discret time is the same as n_a=1.
[4] It is straightforward to add an interaction term to the VARX model and maintain fast OLS estimation.
[5] The original DCM proposed for fMRI included an added complication of modeling the hemodynamic response, which amounts to adding a temporal filter to each output node and prior to adding observation noise.

connectivity, i.e. we ask, can the VARX model recover the underlying structural connectivity, at least in a simulated whole-brian model with known connectivity? We simulated neural activity for a whole-brain neural mass model [45]. We used the default model simulation of the neurolib python library (using their sample code for the “ALNModel”), which is a mean-field approximation of adaptive exponential integrate-and-fire neurons. This model can generate simulated mean firing rates in 80 brain areas based on connectivity and delay matrices determined with diffusion tensor imaging (DTI). We used 5 min of “resting state” activity (no added stimulus, simulated at 0.1ms resolution, subsequently downsampled to 100Hz). The VARX model was estimated with n_a=2, and no input. The resulting estimate for $\mathbf{A}$ is dominated by the diagonal elements that capture the autocorrelation within brain areas (Fig. S1). The true connectivity matrix from DTI (Fig. 2A) is similar to the effect size estimate R for the recurrent connections (Fig. 2B). Following [44] we compare the two as a scatter plot (Fig. 2C) and observed a Spearman correlation of 0.69. For comparison, we also used the sparse-inverse covariance method to recover connectivity from the correlation matrix (functional connectivity). This method is considered state-of-the-art as it is more sensitive than other methods in detecting structural connections [46] and uses the graphical lasso algorithm [47]. The resulting connectivity estimate (Fig. 2D) only achieves a Spearman correlation of 0.52. We note that the structural connectivity determined with DTI is largely symmetric. When enhancing the asymmetry the VARX model is not as accurate, but correctly recovers the direction of the asymmetry (Fig. S2).

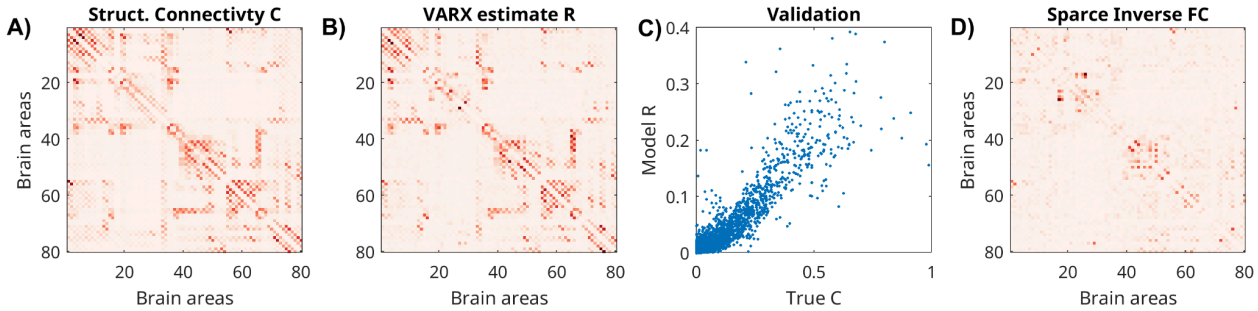

Figure 2: Structural connectivity of stimulated neural mass model for the whole brain, and estimated recurrent connectivity in VARX model. A) True structural connectivity C used to simulate neural activity using a neural mass model with the neurolib python toolbox. Structural connectivity is based on diffusion tensor imaging data between 80 brain areas (called Cmat in neurolib). Here showing the square root of the “Cmat" matrix for better visibility of small connectivity values. B) Effect size estimate R for the recurrent connectivity matrix $\mathbf{A}$ of the VARX model on the simulated data. The diagonal in R is omitted as it is also missing in the structural connectivity Cmat. C) Comparison of true and VARX estimate of connectivity. D) Absolute value of the sparse-inverse functional connectivity (estimated using graphical lasso [48]).

Intracranial EEG recordings and stimulus features

We analyzed intracranial EEG and simultaneous eye-tracking data recorded from patients (N=26 recordings, mean age 38.69 years, age range 19-59 years, 11 female, Table S1) during rest and while they watched various video clips. Four out of 22 individual patients underwent two implantations and recordings at different times resulting in a total of 26 recording sessions with a total of 5,093 recording channels. The video clips included animations with speech

('Despicable Me', two different clips, 10 min each, in English and Hungarian), an animated short film with a mostly visual narrative and music, shown twice ('The Present', 4.3 min), and three clips of documentaries of macaques ('Monkey', 5 min each, without sound) [49]. In addition to the clips from the previous analysis, we included a movie clip of abstract animations ('Inscapes', 10 min) [50], and an eyes-open resting state with maintained fixation ('Resting state', 5 min), and eyes-closed resting state ('Eyes Closed Rest', 5min). In total, we recorded up to 64.7 minutes of data for each patient (Table S1).

Neural signals were preprocessed as previously described to reduce noise [49]. We re-reference signals in a bipolar montage to ensure analysis of local activity. We analyze local field potentials (LFPs) and broadband high-frequency activity (BHA) power. BHA is the power of the signal bandpass filtered between 70-150Hz. We perform analysis on both signals after downsampling to 60Hz. Example traces of y(t) for LFP and BHA are shown in Fig. 1B&C.

We extract six features of the movies that serve as external inputs for the VARX model: fixation onset, fixation novelty, film cuts, motion, sound envelope and acoustic edges (Fig. 1B). Fixation onset, fixation novelty, film cuts and acoustic edges are represented in x(t) as pulse trains with pulses occurring at the time of these events [49]. Motion and the sound envelope are continuous regressors. Motion is the average optic flow across each frame [49]. Fixation novelty is computed as the euclidean distance between features of a convolutional neural network computed on pre-and post fixation image patches. Fixation novelty aims to capture the change of the semantics of visual input across eye movements [49]. The fixation novelty impulses are the same as the fixation regressor, but with their amplitude scaled by novelty. Sound envelope is computed as the absolute value of the Hilbert transform of the sound from the movie files. The envelope is downsampled to 60 Hz. Acoustic edges are peaks in the derivative of the sound envelope, representing rapid changes in the input [51,52]. All videos and resting state include fixations. Since the visual environment is constant during fixations, there is no fixation novelty regressor. The video 'Inscapes', resting state and eyes-closed rest do not include film cuts as external input. The 'Monkey' video clips, resting state, and eyes-closed rest do not include the sound envelope or acoustic edges as input features. Eyes-closed rest does not include any external inputs. When a feature is not available it is replaced with features from a different recording. Therefore, the statistics of the feature are consistent, but not aligned to the neural recording. When comparing models with different features we always keep the number of input variables consistent between models to avoid a bias by the number of free parameters of the model. Features that are not considered in the analysis are shuffled in time by a circular shift by half the duration of the signals.

The models were fitted to data with the matlab version of the publicly-available VARX code [31] using conventional L2-norm regularization. The corresponding regularization parameter was set to λ=0.3. For all analyses we use filters of 600 ms length for inputs (n_b=36 samples for VARX models, L=36 samples for mTRF models). Delays for connections between channels are set to 100ms (n_a=6 samples) for both LFP and BHA signals. Increasing the number of delays n_a, increases estimated effect size R (Fig. S3A,B), however, larger values lead to fewer significant connections (Fig. S3C). Significance (p-value) is computed analytically, i.e. non-parametrically,

based on deviance. Values around n_a=6 time delays appear to be the largest model order supported by this statistical analysis.

Connectivity plots in Fig. 4 were created with routines from the nilearn toolbox [53]. We plot only significant connections (p<0.001). Surface plots of T1w/T2w ratios and directionality of connections are created using the field-echos repository [54,55]. T1w/T2w maps [56] are obtained from the neuromaps repository [57,58], and transformed to the freesurfer surface using code from the neuromaps toolbox [59,60].

The length of responses for each channel in $\mathbf{B}$ and $\mathbf{H}$ to external inputs in Fig. 5 is computed with Matlab's findpeaks() function. This function returns the full-width at half of the peak maximum minus baseline. Power in each channel is computed as the squares of the responses averaged over the time window that was analyzed (0-0.6s).

To compare recurrent connectivity between movies and the resting-state (in Fig. 3), we compute VARX models in four different movie segments of 5 minutes length to match the length of the resting state recording. We use the first and second half of 'Despicable Me English', the first half of 'Inscapes' and one of the 'Monkey' movies. 18 patients include each of these recordings. For each recording in each patient we compute the fraction of significant channels (p<0.001) and average the effect size R across all channel pairs, excluding the diagonal. We test the difference between movies and resting-state with linear mixed-effect models with stimulus as fixed effect (movie vs rest), and patient as random effect (to account for the repeated measures for the different video segments), using matlab's fitlme() routine. For the analysis of asymmetry of recurrent connectivity (in Fig. 4) we also used a mixed-effect model with T1w/T2w ratio as fixed effect and patients as random effect (to account for the repeated measures in multiple brain locations).

Results

Extrinsic input leads to spurious recurrent connectivity

To determine the effect of the extrinsic inputs on connectivity estimates we either fit a VARX model or a VAR model (i.e. a VARX model with no external input). We analyze LFP data on all available recordings, movies and resting state for all N=26 recording sessions. As extrinsic inputs we included film cuts, motion, fixation onset, fixation novelty, the sound envelope and acoustic edges. VAR models contain the same external inputs as the VARX model, but the time alignment is disrupted by a circular shuffle. This keeps the number of parameters in different models constant and ensures the inputs have the same covariance structure. We found a similar connectivity structure for the estimated VAR and VARX models (Fig. 3A and 3B). However, they vary systematically in the number of significant recurrent connections $\mathbf{A}$ (those with p<0.0001, Fig. 3D), which drops when adding inputs (median=-7.3*10^{-4}, p<0.0001, N=26, Wicoxon). The effect sizes R also significantly decreases in the VARX model (Fig. 3E, median=-2.2*10^{-5}, p<0.0001, N=26, Wicoxon). Therefore, accounting for the external input

removes spurious “connections”. We also analyzed how much each of these extrinsic inputs contributed to this effect. Removing any of the input features increased the effect size of recurrent connections compared to a model with all features (Fig. S4). We then cumulatively added each feature to the VARX model. Effect size monotonically decreases with each feature added (Fig. 3F). Decreases of effect size are significant when adding film cuts ($\Delta R=-3.6*10^{-6}$, $p<0.0001$, N=26, FDR correction, $\alpha=0.05$) and the sound envelope ($\Delta R=-3.59*10^{-6}$, $p=0.002$, N=26, FDR correction, $\alpha=0.05$). Thus, adding more input features progressively reduces the strength of recurrent “connections”.

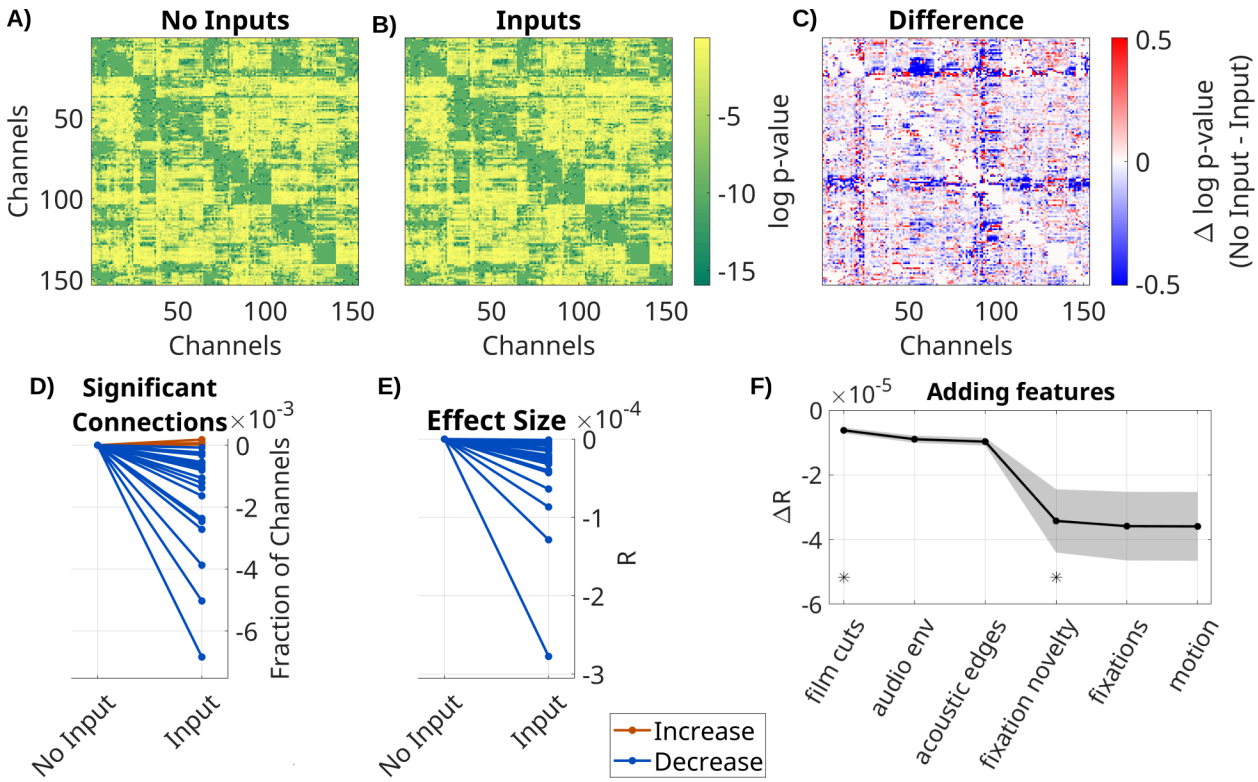

Figure 3: Spurious recurrent connectivity in $\mathbf{A}$ is removed when modeling the effect of extrinsic input with $\mathbf{B}$. Comparison of VARX model with and without inputs. A) log p-values for each connection in $\mathbf{A}$ for a VARX model without inputs on one patient (Pat_1); B) for a VARX model with inputs; C) difference of log p for VAX model without minus with inputs (panel A - B). Both models are fit to the same data. D) Thresholding panels A and B at p<0.0001 gives a fraction of significant connections. Here we show the fraction of significant channels for models with and without input. Each line is a patient with color indicating increase or decrease. E) Mean R over all channels for VARX models with and without inputs. Values in D) and E) have been normalized to models without input. F) Change in R values when successively adding inputs to the VARX model. Black line shows mean across patients, shaded gray area the standard error of the mean. Stars indicate features that further reduce effect size over the previously added feature with statistical significance (Wilcoxon rank sum test p<0.05). Negative values indicate a decrease in connectivity strength when the extrinsic inputs are accounted for. Results for BHA are shown in Fig. S5.

Recurrent connectivity is reduced during movies compared to rest

Next we compared recurrent connectivity between movie watching and rest (Fig. 4). In the rest condition patients have a fixation cross on a gray background. This obviously reduces the size and number of saccades as compared to movie watching, but does not abolish them (Fig. S6). We therefore use a VARX model including fixation onset as extrinsic variable in both cases. Movies include fixation novelty, film cuts, the sound envelope, acoustic edges and motion as external inputs. To control for the number of free parameters, we include copies of features from the movies in the resting state model. The number of significant recurrent connections in $\mathbf{A}$ were significantly reduced during movie watching compared to rest (Fig. 4C, fixed effect of stimulus: beta = $-3.8*10^{-3}$, t(88) = -3.9, p<0.001), as is the effect size R (Fig. 4D, fixed effect of stimulus: beta = $-2.5*10^{-4}$, t(88) = -4.1, p<0.001). While the effect size decreases on average, there is some variation across different brain areas (Fig. 4E-G). In a subset of patients with eyes-closed resting state we find the same effect, that is qualitatively more pronounced (Fig. S8).

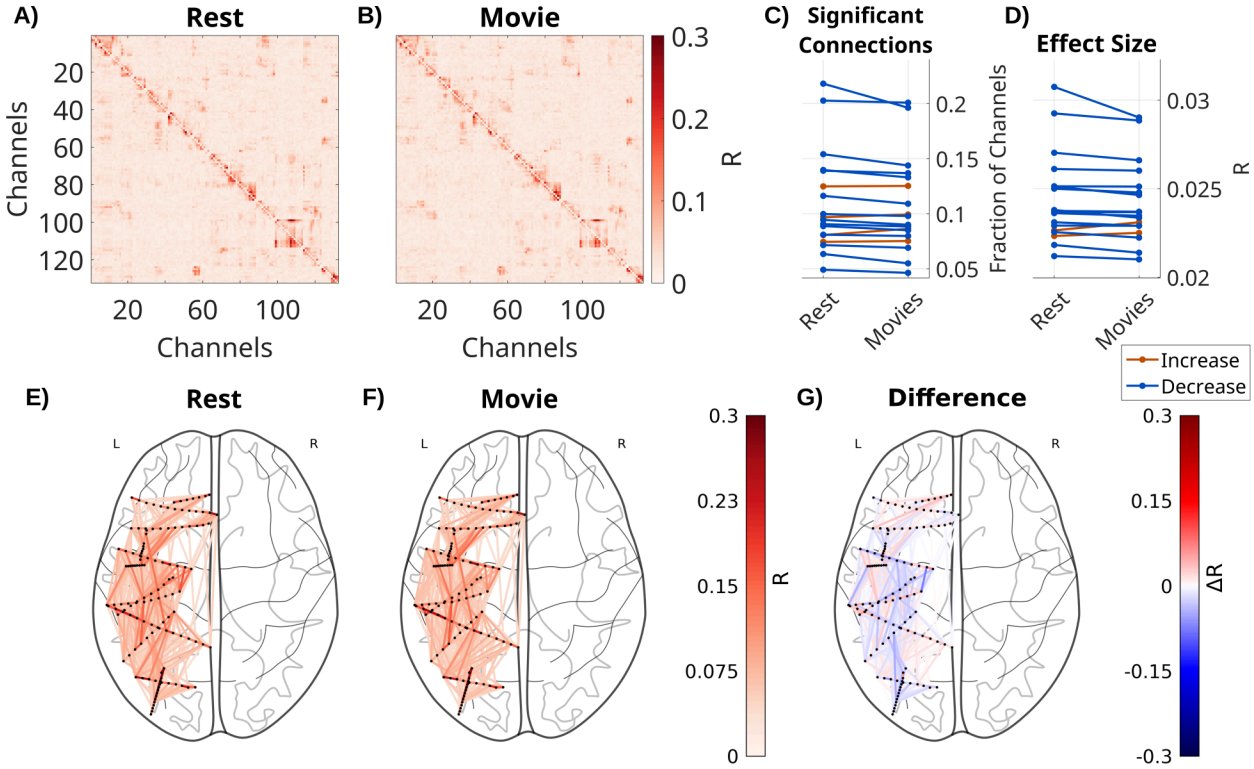

Figure 4: Recurrent connectivity during movies is decreased compared to rest. Effect size R for each connection in $\mathbf{A}$ of one patient (Pat_7) for A) a VARX model during resting fixation with fixation onset as input feature. B) The same with a VARX model of LFP recordings during movie watching, with the following input features: sound envelope, acoustic edges, fixation onsets, fixation novelty, motion and film cuts. C) Fraction of significant connections (p<0.0001)

for movies and rest. D) Mean effect size across all channels for movies and rest. Each line is a patient, with color indicating a numerical increase or decrease. For the movie conditions, we averaged across four different 5 min movie segments. E) Axial view of significant connections in resting state. Black dots show the location of contacts in MNI space. Lines show significant connections between contacts ($p<0.001$) colored in red according to effect size R. For plotting purposes connections in the upper triangle are plotted and asymmetries ignored. F) The same for the movie task, and G) the difference between movies and resting state, showing both increases or decreases for specific connections. Differences between BHA recurrent connectivity $\mathbf{A}$ during movies and resting-state are shown in Figure S7.

Recurrent connectivity enhances and prolongs stimulus responses

We also compared the feed-forward extrinsic effect $\mathbf{B}$ with the total system response $\mathbf{H}$, which includes the additional effect of the recurrent connectivity $\mathbf{A}$. We estimate $\mathbf{B}$ with the VARX model (Fig. 5A) on data during video watching and resting state, and estimate the total response $\mathbf{H}$ directly using temporal response functions (Fig. 5B). Both models include fixation onset, film cuts and sound envelope as external inputs. We compare the power and length of filters from both models (Fig. 5C-D). We compare responses in channels with significant effects of $\mathbf{B}$ (FDR correction, $\alpha=0.05$). We see that the total response $\mathbf{H}$ fixation onset is significantly stronger (Fig. 5C, median$\Delta=1.5*10^{-4}$, $p<0.0001$, N=23, Wilcoxon) and longer than the feed-forward effect $\mathbf{B}$ (Fig. 5D, medianΔ=10.9ms, $p=0.0002$, N=23, Wilcoxon). The same effect is observed for BHA (Fig. S9) and other input features (Fig. S10). This suggests that the total response of the brain to these external inputs is dominated by the intrinsic dynamic of the brain. As with conventional linear regression, the estimate in $\mathbf{B}$ for a particular input and output channel is not affected by which other signals are included in $\mathbf{x}(t)$ or $\mathbf{y}(t)$, provided those other inputs are uncorrelated. We confirmed this here empirically by removing dimensions from $\mathbf{y}(t)$ (Fig. S11A), and by adding uncorrelated input to $\mathbf{x}(t)$ (Fig. S11B, adding fixation onset does not affect the estimate for auditory envelope responses). In other words, to estimate $\mathbf{B}$, we do not require all possible stimulus features and all brain activity to be measured and included in the model. In contrast, $\mathbf{B}$ does vary when correlated inputs are added to $\mathbf{x}(t)$ (Fig. S11C, adding acoustic edges changes the auditory envelope response). Evidently the auditory envelope and acoustic edges are tightly coupled in time, whereas fixation onset is not. When a correlated input is missing (acoustic edges) then the other input (auditory envelope) absorbs the correlated variance, thus capturing the combined response of both.

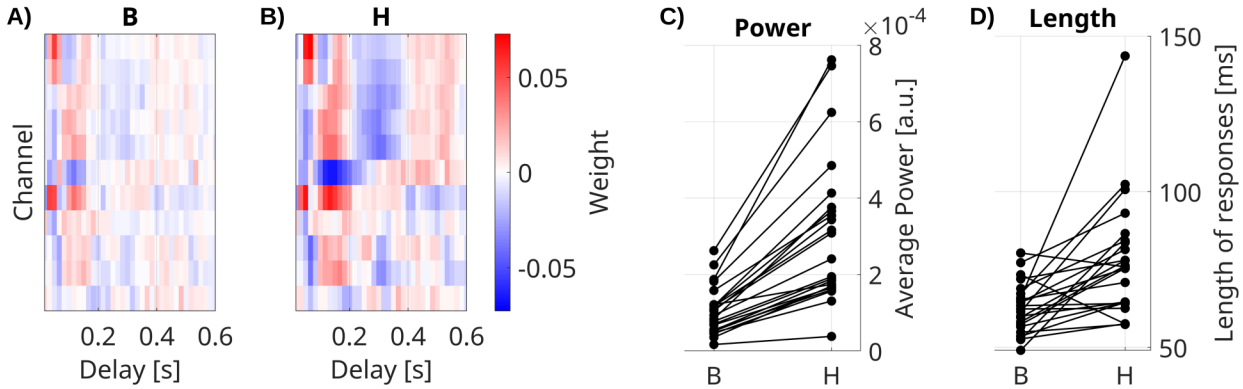

Figure 5: Impulse response models. A) Feed-forward responses $\mathbf{B}$ to fixation onset are weaker and shorter than B) the overall system response $\mathbf{H}$. Impulse response models for fixation onset in channels with significant responses for one example patient. C) Power and D) mean length of responses in significant channels for all patients. Each line is a patient. Responses to fixation onset in all significant channels, as well as auditory envelope and film cuts are shown in Figure S10.

Results are similar for VARX models of BHA and LFP

We repeated the same analyses of Figures 3-5 with broadband high frequency activity (BHA). While LFP are thought to capture dendritic currents, BHA is correlated with neuronal firing rates in the vicinity of an electrode. Generally we find a more sparse recurrent connectivity for BHA as compared to LFP (compare Fig. 3&4 with Fig. S5&S7). Perhaps this is expected, given that LFP covers a broader frequency range. Regardless of this overall difference, we find similar results when analyzing BHA with the VARX model. Namely, taking the extrinsic input into account removed stimulus-induced recurrent connections (Fig. S5); the fraction of significant channels decreases during movie watching compared to rest (Fig. S7); and responses to the stimulus are stronger and more prolonged when separately modeling the effect of recurrent connectivity (Fig. S9). In the Discussion section we will argue that some of these results are expected in general when decomposing the total system response into extrinsic and intrinsic effects.

Intrinsic “noise” in BHA is reduced by external stimulus

So far we have discussed the mean response captured by $\mathbf{B}$ and the activity mediated by $\mathbf{A}$. We now want to analyze whether the external input modulates the internal variability of brain activity. As a metric of internal variability we measured the power of the intrinsic innovation process $\mathbf{e}(t)$, which captures the unobserved “random” brain activity which leads to variations in the responses. For the LFP signal we see a drop in power during movies as compared to rest, for both the original signal $\mathbf{y}(t)$ (Fig. S12A) and the model’s innovation process $\mathbf{e}(t)$ (Fig. S12B). Notable is the stronger oscillatory activity during rest (Fig. S12A). In this example we see a drop in power in the theta/alpha band (5-11 Hz) during movie watching across all electrodes (Fig. S12A, dotted lines). We observe similar narrow-band drop in power in most patients, albeit at different frequencies (not shown). When analyzing BHA, we find no difference in power of the

innovation process between movie and rest, but we do find a drop in power relative to the overall BHA signals for some channels (Fig. 6B). These channels seem to coincide with channels that responded to the external stimuli, i.e. channels with a significant effect in $\mathbf{B}$ (Fig. 6A). If we take for each patient the median relative power for responsive channels (median among those with p<0.0001), then we find that relative power drops for nearly all patients (Fig 6D, Wilcoxon rank sum test, $p=8.8*10^{-7}$, N=26). The motivation for analyzing only responsive channels comes from a simple gain adaptation (Fig. S13). Gain adaptation keeps the power of $\mathbf{y}(t)$ constant, so that the extra power injected by the stimulus implicitly reduces the relative power of the innovation process. This effect is specific to channels receiving external input (Fig. S13D) and absent in a linear system without gain adaptation (Fig. S13C). To demonstrate that this simple gain adaptation can explain the noise quenching in the neural data, we simulated data with the gain adaptation model (Fig. 6C) using parameters estimated for the example patient of Fig. 6A/B.

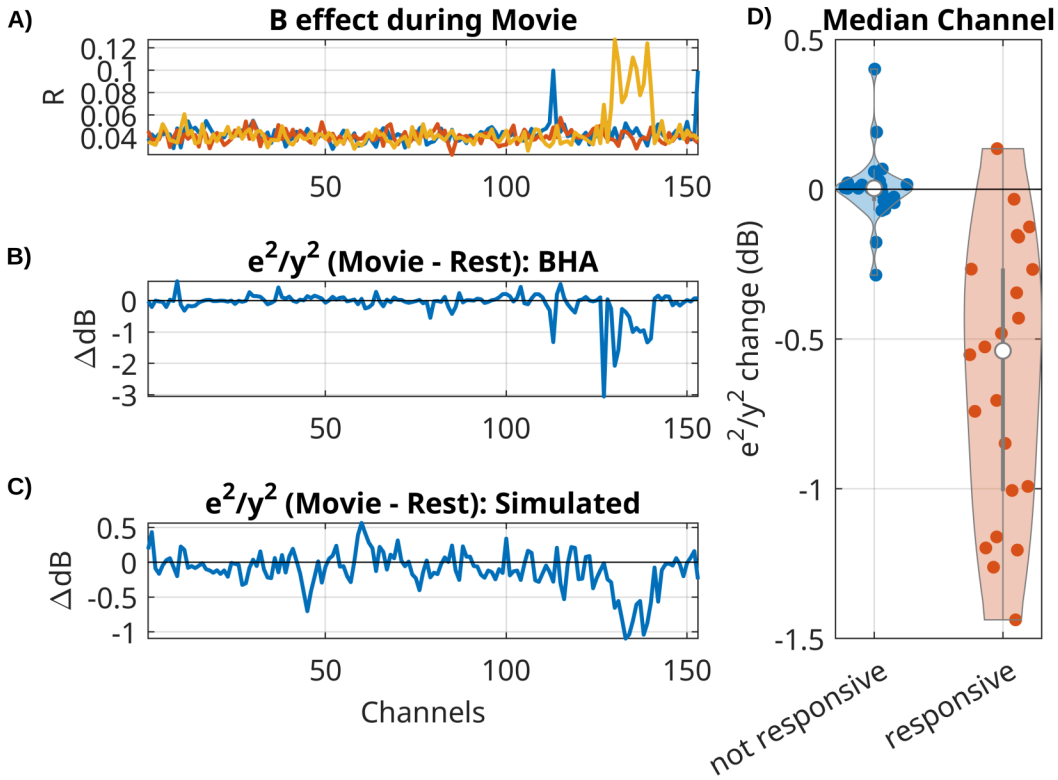

Figure 6: For BHA, relative power of innovation vs signal drops during movies as compared to rest in responsive channels. A) Effect size R for extrinsic effect $\mathbf{B}$ in all channels for 3 input features (scene cuts, fixation onset, sound envelope). In this example 15 electrodes had significant responses to one of the three inputs (Bonferroni corrected at p<0.01). B) Change in relative power of innovation (dB(innovation power / signal power), then subtracting movie - rest). C) Change in relative power of innovation in a simulation of a VARX model with gain adaptation. Here we are using the $\mathbf{A}$ and $\mathbf{B}$ filters that were estimated on BHA on the example from panel A and B. D) Median of power ratio change across all patients, contrasting responsive vs non-responsive channels.

Direction of connectivity differs with cortical hierarchy

Finally, we measured the directionality of the recurrent connections in the LFPs by analyzing the structure of the resulting matrices R of all patients, combining data from all available movies and resting state recordings. Columns in R represent outgoing connections, while rows are incoming connections. Therefore, the difference of $R - R^\top$ (Fig. 7A) averaged along a column has positive values if a node has overall stronger outgoing connections, and negative values if it has stronger incoming connections. We measured this directionality for each channel across all patients and averaged also across channels within parcels of the Desikan-Killiany atlas (N=34 regions of interest, Fig. 7B) [61]. We expected this to co-vary with "cortical hierarchy". To test this, we compared this asymmetry metric with the T1w/T2w ratio, which captures gray matter myelination and is used as an indirect measure of cortical hierarchy [54,62]. We also average T1w/T2w ratio in the same parcels of the Desikan-Killiany atlas (Fig. 7B). We used a mixed effect model and find that cortical areas showing more outgoing connections ($R - R^\top > 0$) have higher T1w/T2w ratio, which are located lower on the cortical hierarchy (t(533)=2.62, p = 0.009, Fig. 7C). BHA analysis shows the same effect (Fig. S14).

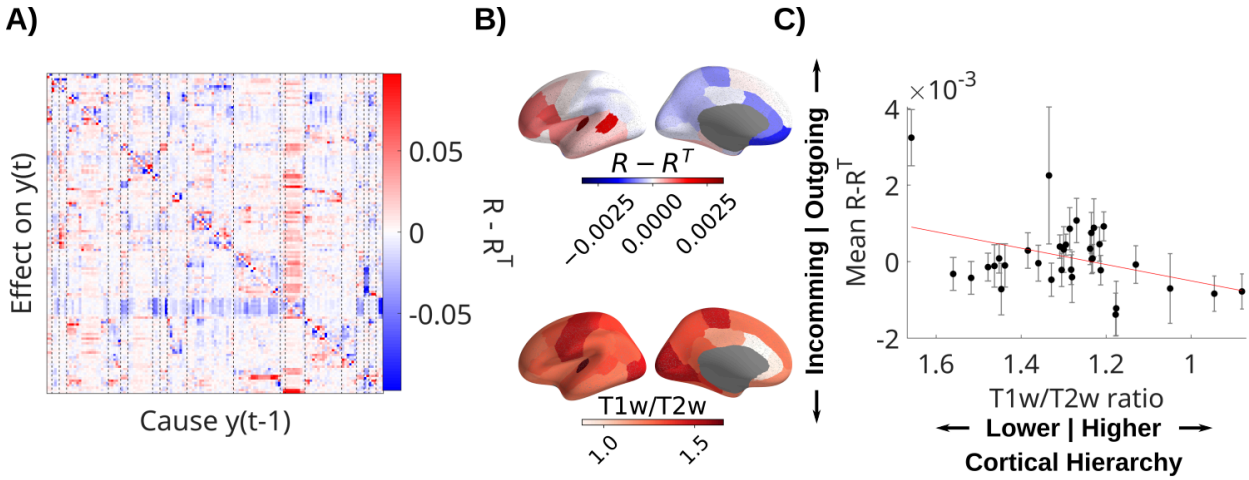

Figure 7: Recurrent connectivity of LFP is directed from sensory to higher-order areas. A) Difference of $R - R^\top$ showing asymmetric directed effects. Dashed lines indicate regions of interest in the Desikan-Killiany atlas. B) Mean directionality across patients and T1w/T2w ratio are averaged in parcels of the Desikan-Killiany atlas. C) Mean directionality is correlated with cortical hierarchy, estimated with the T1w/T2w ratio. Each dot represents a parcel in the Desikan-Kiliany atlas with error-bars indicating error of the mean across patients with channels in that parcel. The datapoint on the top left is the transverse temporal gyrus. Note that the x-axis has been flipped to show areas higher on the cortical hierarchy on the right. T1w/T2w ratio and cortical hierarchy have an inverse relationship.

Discussion

Our results suggest that the duration and magnitude of responses to extrinsic input is in large part a result of the intrinsic dynamic of the recurrent brain network. We also found that the intrinsic dynamic had reduced recurrent connectivity and weaker intrinsic variability during movie stimulus.

Response to extrinsic input versus intrinsic dynamics

Previous literature does often not distinguish between intrinsic dynamics and extrinsic effects. By factoring out some of the linear effects of the external input we conclude here that recurrent connectivity is reduced on average. From our prior work[49], we know that the stimulus features we included here capture a substantial amount of variance across the brain in intracranial EEG. Arguably, however, the video stimuli had rich semantic information that was not captured by the low-level features used here. Adding such semantic features could have further reduced shared variance, and consequently further reduced average recurrent connectivity in the model.

Similarities and differences between rest and movie watching conditions reported previously, do not draw a firm conclusion as to whether overall “functional connectivity” is increased or reduced. Results seem to depend on the time scale of neural activity analyzed, and the specific brain networks [12,16,63]. However, in fMRI, the conclusion seems to be that functional connectivity during movies is stronger than during rest [15], which likely results from stimulus induced correlations. The VARX model can remove some of the effects of these stimuli, revealing that average recurrent connectivity may be reduced rather than increased during stimulus processing. Reduced functional connectivity has previously been observed within the visual cortex when a visual stimulus is presented [27].

In this work we focused on "passive" tasks, i.e. resting with gaze on a fixation point, versus watching movies without any associated tasks. We did not analyze data during an active task requiring behavioral responses. The literature on active tasks emphasizes "state change" in functional connectivity. [14,20,64] Efforts to factor out task-evoked activity when computing functional connectivity concord with our conclusions that connectivity is inflated by a task [17]. Nevertheless, we hesitate extrapolating our findings to active tasks, as we have not analyzed such data. Future studies should test if our findings replicate in an independent iEEG datasets, including active tasks and whether they generalize to other neuroimaging modalities.

Conventional “encoding” models, such as temporal response functions, capture the total response $\mathbf{H}$ of the brain to an external stimulus. Here we factored this into a moving average filter $\mathbf{B}$, followed by and autoregressive filter $\mathbf{A}$. The important observation is that this intrinsic dynamic governed by $\mathbf{A}$ does not change during stimulus processing. Arguably then, the role of the initial responses $\mathbf{B}$ is to shape the input to be processed by the existing intrinsic dynamic. This interpretation is consistent with the view of “the brain from the inside out” advocated by György Buzsáki [65]. In this view, learning of a stimulus representation consists in learning a mapping of the external stimulus to an existing intrinsic dynamic of the brain.

Similar findings for LFP and BHA

We found a more sparse recurrent connectivity for BHA as compared to LFP. This may be expected because correlations in lower frequencies (that dominate LFPs) reach over longer distances compared to correlations in higher frequencies (e.g. [66]). BHA has been linked to a mixture of neuronal firing and dendritic currents [67], in contrast to LFP, which is thought to originate from widespread dendritic currents. Despite the observed differences in sparsity, for both LFP and BHA we found that modeling the intrinsic dynamic removed spurious recurrent connections. Removal of spurious effects when controlling for a common cause is a generic finding in multivariate statistical models. We also found for both LFP and BHA that the duration and strength of stimulus responses can be largely attributed to the recurrent connections. Arguably, this is a generic feature of an autoregressive model, as it more readily captures longer impulse responses. However, the extrinsic filters $\mathbf{B}$ in principle have an advantage as they can be fit to each stimulus and brain location. In contrast, the recurrent filters $\mathbf{A}$ are constrained by having to capture a shared dynamic for all stimulus dimensions. Thus, the predominance of the intrinsic dynamic in the total system response is not a trivial result of the factorization into intrinsic and extrinsic effects.

Stimulus-induced reduction of noise in the intrinsic activity

One difference we did find between LFP and BHA is the intrinsic innovation process, i.e. the internal sources of variability or “noise”. For both BHA and LFP we saw a drop in the magnitude of signal fluctuations during the movie watching condition. For the BHA but not the LFP, this was explained as a drop in intrinsic noise. Specifically, for BHA there was less relative power in the intrinsic “noise” for channels that are responsive to the stimulus. This is consistent with the notion that response variability is due to variability of intrinsic activity [22] which is found to decrease across the brain with the onset of an external stimulus [68]. This type of noise quenching has been associated with increased attention [69] and improved visual discrimination performance [70]. The effect we found here can be explained by a VARX model with the addition of a divisive gain adaptation mechanism that keeps the total power of brain activity constant. When the input injects additional power, this nonlinear gain adaptation implicitly reduces the contribution of the intrinsic noise to the total power. The noise-quenching result and its explanation via gain adaptation shows the benefit of using a parsimonious linear model, which can suggest nonlinear mechanisms as simple corrections from linearity.

We also observed an overall drop in LFP power during movie watching. This phenomenon was strongest in oscillatory bands, with frequencies in theta (5-8Hz) to beta (15-25Hz) band differing across patients. In scalp EEG, noise quenching is associated with a similar overall drop in power with the stimulus [69]. This quenching of neural variability was also found to reduce correlation between brian areas for fMRI and neural spiking [26]. Both fMRI and neural spiking correlated with BHA [71].

Stimulus features

During the movie and rest periods, we utilized fixation onset to capture activity that is time-locked to visual processing because patients move their eyes even during rest. We also added the fixation novelty regressor to capture semantic changes in the visual input across eye movements [49]. We incorporated the sound envelope, a prominent feature known for capturing the dominant audio-induced variance in scalp EEG [33], as well as acoustic edges that capture strong transients in the auditory input [51,52]. In addition, we included film cuts as features, as we had previously demonstrated that they dominate the response in the BHA across the brain [49]. Motion is added as an additional feature to capture while other basic visual features such as overall optic flow or fixations on faces elicited responses in the BHA, their contribution was relatively smaller. The analysis is not limited to these few features, and future research should explore which stimulus features capture variance in the data and how they affect the apparent recurrent connectivity. There is a substantial body of literature on encoding models of semantic features, where nonlinear features of a continuous natural stimulus are extracted and then linearly regressed against fMRI [72,73] or EEG [74]. This work can be directly replicated with the VARX model which further models the recurrent connectivity.

Alternative approaches

The traditional VAR model has been used extensively in neuroscience to establish directed "Granger causal" connections [41]. This approach has been very fruitful and found numerous extensions, e.g. [10,11]. However, these model implementations do not specifically account for an external input.

A few methods have attempted to model the effect of varying task conditions on functional connectivity, mostly in the analysis of fMRI. One approach is to first model the task-evoked responses, equivalent to estimating $\mathbf{B}$ alone, and then compute the conventional "functional connectivity", i.e. the correlation matrix, on the residuals $\boldsymbol{e}(t)$ [75]. Others suggested to estimate $\mathbf{B}$ in multiple time windows and then estimate a "task related functional connectivity" by correlating the multiple $\mathbf{B}$ over time windows [76]. It is not clear that these ad-hoc methods systematically separate intrinsic from extrinsic factors.

A more principled modeling approach is "dynamic causal modeling" (DCM) [19] and extensions thereof [77]. Similar to the VARX model, DCM includes intrinsic and extrinsic effects $\boldsymbol{A}$ and $\mathbf{B}$. However, the modeling is limited to first-order dynamics (i.e. n_a=n_b=1). Thus, prolonged responses have to be entirely captured with a first-order recurrent $\mathbf{A}$. In contrast, the DCM includes a multiplicative interaction of extrinsic input $\boldsymbol{x}(t)$ on the connectivity $\mathbf{A}$, which does not exist in the VARX model. This interaction has been used to explicitly model a change in recurrent connectivity with task conditions. Here we found that this may not be necessary for intracranial EEG. A practical advantage of the VARX model is the assumption that the neural activity is directly observed. Instead, many existing models assume an error in the observations, which triggers computationally intensive estimation algorithms, typically the expectation maximization algorithm. The same is true for the "output error" model in linear systems theory [29].

As a result, these models are often limited to small networks[6] to test specific alternative hypotheses [42]. In contrast, here we have analyzed up to 300 channels per patient across the brain, which would be prohibitive with DCM. By analyzing a large number of recordings we were able to draw more general conclusions about whole-brain activity.

Caveats:

The stimulus features we included in our model capture mostly low-level visual and auditory input. It is possible that regressing out a richer stimulus characterization would have removed additional stimulus-induced correlation. While we do not expect that this would change the overall effect of a reduced number of “connections” during movie watching compared to resting state, the interpretation of changes in specific connections will be affected by the choice of features. For example, in sensory cortices, higher recurrent connectivity in the LFP during rest would be consistent with the more synchronized state we saw in rest, as reflected by larger oscillatory activity. Synchronization in higher-order cortices, however, is expected to be more strongly influenced by semantic content of external input.

We find a correlation of DTI structural connectivity used in a model with a VARX estimate of 0.70. That is considered a relatively large value compared to other studies that attempt to recover DTI connectivity from the correlation structure of fMRI activity [44]. A Caveat is that this was done on a biophysical model of firing rate, not fMRI, and we have not explored the parameters of the model that might affect the results.

We used fixation onsets as external input, but it should be noted that they are tightly correlated in time with saccade onsets (there is only about a 30 ms jitter between the two, depending on saccade amplitude). While saccades are driven by visual movement, they are generated by the brain itself and arguably could also be seen as intrinsic. The same is true for all motor behaviors, most of which cause a corresponding sensory response, similar to the visual response following a saccade. Including them as external input is a modeling choice we have made here, but it is important to acknowledge that fixation onsets can therefore have “acausal” components [49]. By “acausal” we mean a fixation-locked response that precedes the fixation onset and is due to the neural activity leading up to the saccade and subsequent fixation. Such acausal responses can be captured by the VARX Granger formalism by delaying the input relative to the neural activity, which we have not done here.

The correlation between the average incoming and outgoing connections and cortical hierarchy (Fig. 7) is not significant when normalizing for the number of electrodes in each region of interest. Regions in the temporal lobe with a large number of electrodes might drive this correlation. A more fine grained analysis in these regions could be the goal of future analysis.

---

[6] The original DCM proposed for fMRI included an added complication of modeling the hemodynamic response, which amounts to adding a temporal filter to each output node and prior to adding observation noise.

Conclusion

We analyzed whole-brain intracranial recordings in human patients at rest and while they watched videos. We used a model that separates intrinsic dynamics from extrinsic effects. We found that the brain's response to the audiovisual stimuli appears to be substantially shaped by its endogenous dynamic. The model revealed a small but significant decrease of recurrent connectivity when watching movies. Finally, we observed a reduction in intrinsic variance during the extrinsic stimulus, which may be the result of neuronal gain adaptation.

Acknowledgements

We would like to thank Chris Honey for advice on the model validation with simulations and related references. We like to thank Behtash Babadi for help on the development of the Granger formalism for the VARX model. This work was supported in part by the NIH through grants P50 MH109429 and R01DC019979.

References

1. Felleman, D.J., and Van Essen, D.C. (1991). Distributed Hierarchical Processing in the Primate Cerebral Cortex. Cereb. Cortex *1*, 1–47. https://doi.org/10.1093/cercor/1.1.1-a.

2. Friston, K.J., Holmes, A.P., Poline, J.-B., Grasby, P.J., Williams, S.C.R., Frackowiak, R.S.J., and Turner, R. (1995). Analysis of fMRI Time-Series Revisited. NeuroImage *2*, 45–53. https://doi.org/10.1006/nimg.1995.1007.

3. Lalor, E.C., and Foxe, J.J. (2010). Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution. Eur. J. Neurosci. *31*, 189–193. https://doi.org/10.1111/j.1460-9568.2009.07055.x.

4. Greicius, M.D., Krasnow, B., Reiss, A.L., and Menon, V. (2003). Functional connectivity in the resting brain: A network analysis of the default mode hypothesis. Proc. Natl. Acad. Sci. *100*, 253–258. https://doi.org/10.1073/pnas.0135058100.

5. Varela, F., Lachaux, J.-P., Rodriguez, E., and Martinerie, J. (2001). The brainweb: Phase synchronization and large-scale integration. Nat. Rev. Neurosci. *2*, 229–239. https://doi.org/10.1038/35067550.

6. Friston, K., Moran, R., and Seth, A.K. (2013). Analysing connectivity with Granger causality and dynamic causal modelling. Curr. Opin. Neurobiol. *23*, 172–178. https://doi.org/10.1016/j.conb.2012.11.010.

7. Haufe, S., Nikulin, V.V., Müller, K.-R., and Nolte, G. (2013). A critical assessment of connectivity measures for EEG data: A simulation study. NeuroImage *64*, 120–133. https://doi.org/10.1016/j.neuroimage.2012.09.036.

8. Pellegrini, F., Delorme, A., Nikulin, V., and Haufe, S. (2023). Identifying good practices for

detecting inter-regional linear functional connectivity from EEG. NeuroImage *277*, 120218. https://doi.org/10.1016/j.neuroimage.2023.120218.

9. Seth, A.K., Barrett, A.B., and Barnett, L. (2015). Granger Causality Analysis in Neuroscience and Neuroimaging. J. Neurosci. *35*, 3293–3297. https://doi.org/10.1523/JNEUROSCI.4399-14.2015.

10. Sheikhattar, A., Miran, S., Liu, J., Fritz, J.B., Shamma, S.A., Kanold, P.O., and Babadi, B. (2018). Extracting neuronal functional network dynamics via adaptive Granger causality analysis. Proc. Natl. Acad. Sci. *115*, E3869–E3878. https://doi.org/10.1073/pnas.1718154115.

11. Soleimani, B., Das, P., Dushyanthi Karunathilake, I.M., Kuchinsky, S.E., Simon, J.Z., and Babadi, B. (2022). NLGC: Network localized Granger causality with application to MEG directional functional connectivity analysis. NeuroImage *260*, 119496. https://doi.org/10.1016/j.neuroimage.2022.119496.

12. Betti, V., Della Penna, S., de Pasquale, F., Mantini, D., Marzetti, L., Romani, G.L., and Corbetta, M. (2013). Natural Scenes Viewing Alters the Dynamics of Functional Connectivity in the Human Brain. Neuron *79*, 782–797. https://doi.org/10.1016/j.neuron.2013.06.022.

13. Geerligs, L., Rubinov, M., Cam-CAN, and Henson, R.N. (2015). State and Trait Components of Functional Connectivity: Individual Differences Vary with Mental State. J. Neurosci. *35*, 13949–13961. https://doi.org/10.1523/JNEUROSCI.1324-15.2015.

14. Mennes, M., Kelly, C., Colcombe, S., Castellanos, F.X., and Milham, M.P. (2013). The Extrinsic and Intrinsic Functional Architectures of the Human Brain Are Not Equivalent. Cereb. Cortex *23*, 223–229. https://doi.org/10.1093/cercor/bhs010.

15. Vanderwal, T., Eilbott, J., Finn, E.S., Craddock, R.C., Turnbull, A., and Castellanos, F.X. (2017). Individual differences in functional connectivity during naturalistic viewing conditions. NeuroImage *157*, 521–530. https://doi.org/10.1016/j.neuroimage.2017.06.027.

16. Demirtaş, M., Ponce-Alvarez, A., Gilson, M., Hagmann, P., Mantini, D., Betti, V., Romani, G.L., Friston, K., Corbetta, M., and Deco, G. (2019). Distinct modes of functional connectivity induced by movie-watching. NeuroImage *184*, 335–348. https://doi.org/10.1016/j.neuroimage.2018.09.042.

17. Cole, M.W., Ito, T., Schultz, D., Mill, R., Chen, R., and Cocuzza, C. (2019). Task activations produce spurious but systematic inflation of task functional connectivity estimates. NeuroImage *189*, 1–18. https://doi.org/10.1016/j.neuroimage.2018.12.054.

18. Pearl, J. (2013). Linear Models: A Useful “Microscope” for Causal Analysis. J. Causal Inference *1*, 155–170. https://doi.org/10.1515/jci-2013-0003.

19. Friston, K.J., Harrison, L., and Penny, W. (2003). Dynamic causal modelling. NeuroImage *19*, 1273–1302. https://doi.org/10.1016/S1053-8119(03)00202-7.

20. Gonzalez-Castillo, J., and Bandettini, P.A. (2018). Task-based dynamic functional connectivity: Recent findings and open questions. NeuroImage *180*, 526–533. https://doi.org/10.1016/j.neuroimage.2017.08.006.

21. Fox, M.D., Snyder, A.Z., Zacks, J.M., and Raichle, M.E. (2006). Coherent spontaneous activity accounts for trial-to-trial variability in human evoked brain responses. Nat. Neurosci. *9*, 23–25. https://doi.org/10.1038/nn1616.

22. Arieli, A., Sterkin, A., Grinvald, A., and Aertsen, A. (1996). Dynamics of Ongoing Activity: Explanation of the Large Variability in Evoked Cortical Responses. Science *273*, 1868–1871. https://doi.org/10.1126/science.273.5283.1868.

23. Buzsaki, G. (2006). Rhythms of the Brain (Oxford University Press).

24. Fiser, J., Chiu, C., and Weliky, M. (2004). Small modulation of ongoing cortical dynamics by sensory input during natural vision. Nature *431*, 573–578. https://doi.org/10.1038/nature02907.

25. Gray, C.M., König, P., Engel, A.K., and Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. Nature *338*, 334–337. https://doi.org/10.1038/338334a0.

26. Ito, T., Brincat, S.L., Siegel, M., Mill, R.D., He, B.J., Miller, E.K., Rotstein, H.G., and Cole, M.W. (2020). Task-evoked activity quenches neural correlations and variability across cortical areas. PLOS Comput. Biol. *16*, e1007983. https://doi.org/10.1371/journal.pcbi.1007983.

27. Nauhaus, I., Busse, L., Carandini, M., and Ringach, D.L. (2009). Stimulus contrast modulates functional connectivity in visual cortex. Nat. Neurosci. *12*, 70–76. https://doi.org/10.1038/nn.2232.

28. Roelfsema, P.R., Engel, A.K., König, P., and Singer, W. (1997). Visuomotor integration is associated with zero time-lag synchronization among cortical areas. Nature *385*, 157–161. https://doi.org/10.1038/385157a0.

29. Ljung, L. (1999). System Identification: Theory for the User (Prentice Hall PTR).

30. Hamilton, J.D. (2020). Time Series Analysis (Princeton University Press).

31. Parra, L.C., Silvan, A., Nentwich, M., Madsen, J., Parra, V.E., and Babadi, B. (2025). VARX Granger analysis: Models for neuroscience, physiology, sociology and econometrics. PLOS ONE *20*, e0313875. https://doi.org/10.1371/journal.pone.0313875.

32. Naselaris, T., Kay, K.N., Nishimoto, S., and Gallant, J.L. (2011). Encoding and decoding in fMRI. NeuroImage *56*, 400–410. https://doi.org/10.1016/j.neuroimage.2010.07.073.

33. Di Liberto, G.M., O'Sullivan, J.A., and Lalor, E.C. (2015). Low-Frequency Cortical Entrainment to Speech Reflects Phoneme-Level Processing. Curr. Biol. *25*, 2457–2465. https://doi.org/10.1016/j.cub.2015.08.030.

34. Holdgraf, C.R., Rieger, J.W., Micheli, C., Martin, S., Knight, R.T., and Theunissen, F.E. (2017). Encoding and Decoding Models in Cognitive Electrophysiology. Front. Syst. Neurosci. *11*. https://doi.org/10.3389/fnsys.2017.00061.

35. Li, Y., Anumanchipalli, G.K., Mohamed, A., Chen, P., Carney, L.H., Lu, J., Wu, J., and

Chang, E.F. (2023). Dissecting neural computations in the human auditory pathway using deep neural networks for speech. Nat. Neurosci., 1–13. https://doi.org/10.1038/s41593-023-01468-4.

36. Granger, C.W.J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. Econometrica *37*, 424–438. https://doi.org/10.2307/1912791.

37. Geweke, J. (1982). Measurement of Linear Dependence and Feedback between Multiple Time Series. J. Am. Stat. Assoc. *77*, 304–313. https://doi.org/10.1080/01621459.1982.10477803.

38. Magee, L. (1990). R 2 Measures Based on Wald and Likelihood Ratio Joint Significance Tests. Am. Stat. *44*, 250–253. https://doi.org/10.1080/00031305.1990.10475731.

39. Crosse, M.J., Di Liberto, G.M., Bednar, A., and Lalor, E.C. (2016). The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli. Front. Hum. Neurosci. *10*. https://doi.org/10.3389/fnhum.2016.00604.

40. Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J. -P., Frith, C.D., and Frackowiak, R.S.J. (1994). Statistical parametric maps in functional imaging: A general linear approach. Hum. Brain Mapp. *2*, 189–210. https://doi.org/10.1002/hbm.460020402.

41. Barnett, L., and Seth, A.K. (2014). The MVGC multivariate Granger causality toolbox: A new approach to Granger-causal inference. J. Neurosci. Methods *223*, 50–68. https://doi.org/10.1016/j.jneumeth.2013.10.018.

42. Penny, W.D., Stephan, K.E., Mechelli, A., and Friston, K.J. (2004). Comparing dynamic causal models. NeuroImage *22*, 1157–1172. https://doi.org/10.1016/j.neuroimage.2004.03.026.

43. Bassett, D.S., Zurn, P., and Gold, J.I. (2018). On the nature and use of models in network neuroscience. Nat. Rev. Neurosci. *19*, 566–578. https://doi.org/10.1038/s41583-018-0038-8.

44. Honey, C.J., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J.P., Meuli, R., and Hagmann, P. (2009). Predicting human resting-state functional connectivity from structural connectivity. Proc. Natl. Acad. Sci. *106*, 2035–2040. https://doi.org/10.1073/pnas.0811168106.

45. Cakan, C., Jajcay, N., and Obermayer, K. (2023). neurolib: A Simulation Framework for Whole-Brain Neural Mass Modeling. Cogn. Comput. *15*, 1132–1152. https://doi.org/10.1007/s12559-021-09931-9.

46. Smith, S.M., Miller, K.L., Salimi-Khorshidi, G., Webster, M., Beckmann, C.F., Nichols, T.E., Ramsey, J.D., and Woolrich, M.W. (2011). Network modelling methods for FMRI. NeuroImage *54*, 875–891. https://doi.org/10.1016/j.neuroimage.2010.08.063.

47. Chen, X. (2023). xiaohuichen88/Graphical-Lasso.

48. Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. Biostatistics *9*, 432–441. https://doi.org/10.1093/biostatistics/kxm045.

49. Nentwich, M., Leszczynski, M., Russ, B.E., Hirsch, L., Markowitz, N., Sapru, K., Schroeder, C.E., Mehta, A.D., Bickel, S., and Parra, L.C. (2023). Semantic novelty modulates neural responses to visual change across the human brain. Nat. Commun. *14*, 2910. https://doi.org/10.1038/s41467-023-38576-5.

50. Vanderwal, T., Kelly, C., Eilbott, J., Mayes, L.C., and Castellanos, F.X. (2015). Inscapes : A movie paradigm to improve compliance in functional magnetic resonance imaging. NeuroImage *122*, 222–232. https://doi.org/10.1016/j.neuroimage.2015.07.069.

51. Forseth, K.J., Hickok, G., Rollo, P.S., and Tandon, N. (2020). Language prediction mechanisms in human auditory cortex. Nat. Commun. *11*, 5240. https://doi.org/10.1038/s41467-020-19010-6.

52. Oganian, Y., and Chang, E.F. (2019). A speech envelope landmark for syllable encoding in human superior temporal gyrus. Sci. Adv. https://doi.org/10.1126/sciadv.aay6279.

53. Chamma, A., Frau-Pascual, A., Rothberg, A., Abadie, A., Abraham, A., Gramfort, A., Savio, A., Cionca, A., Thual, A., Kodibagkar, A., et al. (2024). nilearn. Version 0.10.4 (Zenodo). https://doi.org/10.5281/zenodo.10948303 https://doi.org/10.5281/zenodo.10948303.

54. Gao, R., van den Brink, R.L., Pfeffer, T., and Voytek, B. (2020). Neuronal timescales are functionally dynamic and shaped by cortical microarchitecture. eLife *9*, e61277. https://doi.org/10.7554/eLife.61277.

55. Gao, R., Voytek, B., and Olayinka, T. (2020). rdgao/field-echos: post-publication. Version v1.0 (Zenodo). https://doi.org/10.5281/zenodo.4362645 https://doi.org/10.5281/zenodo.4362645.

56. Glasser, M.F., Coalson, T.S., Robinson, E.C., Hacker, C.D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C.F., Jenkinson, M., et al. (2016). A multi-modal parcellation of human cerebral cortex. Nature *536*, 171–178. https://doi.org/10.1038/nature18933.

57. Markello, R., Hansen, J., Liu, Z.-Q., Bazinet, V., Shafiei, G., Suarez, L., and Mišić, B. (2024). neuromaps: structural and functional interpretation of brain maps. Version 0.0.5 (Zenodo). https://doi.org/10.5281/zenodo.10607923 https://doi.org/10.5281/zenodo.10607923.

58. Markello, R.D., Hansen, J.Y., Liu, Z.-Q., Bazinet, V., Shafiei, G., Suárez, L.E., Blostein, N., Seidlitz, J., Baillet, S., Satterthwaite, T.D., et al. (2022). neuromaps: structural and functional interpretation of brain maps. Nat. Methods *19*, 1472–1479. https://doi.org/10.1038/s41592-022-01625-w.

59. Robinson, E.C., Garcia, K., Glasser, M.F., Chen, Z., Coalson, T.S., Makropoulos, A., Bozek, J., Wright, R., Schuh, A., Webster, M., et al. (2018). Multimodal surface matching with higher-order smoothness constraints. NeuroImage *167*, 453–465. https://doi.org/10.1016/j.neuroimage.2017.10.037.

60. Robinson, E.C., Jbabdi, S., Glasser, M.F., Andersson, J., Burgess, G.C., Harms, M.P., Smith, S.M., Van Essen, D.C., and Jenkinson, M. (2014). MSM: A new flexible framework for Multimodal Surface Matching. NeuroImage *100*, 414–426. https://doi.org/10.1016/j.neuroimage.2014.05.069.

61. Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. NeuroImage *31*, 968–980. https://doi.org/10.1016/j.neuroimage.2006.01.021.

62. Wang, X.-J. (2020). Macroscopic gradients of synaptic excitation and inhibition in the neocortex. Nat. Rev. Neurosci. *21*, 169–178. https://doi.org/10.1038/s41583-020-0262-x.

63. Vanderwal, T., Kelly, C., Eilbott, J., Mayes, L.C., and Castellanos, F.X. (2015). *Inscapes*: A movie paradigm to improve compliance in functional magnetic resonance imaging. NeuroImage *122*, 222–232. https://doi.org/10.1016/j.neuroimage.2015.07.069.

64. Cole, M.W., Bassett, D.S., Power, J.D., Braver, T.S., and Petersen, S.E. (2014). Intrinsic and Task-Evoked Network Architectures of the Human Brain. Neuron *83*, 238–251. https://doi.org/10.1016/j.neuron.2014.05.014.

65. Buzsaki, G. (2019). The Brain from Inside Out (Oxford University Press).

66. Muller, L., Hamilton, L.S., Edwards, E., Bouchard, K.E., and Chang, E.F. (2016). Spatial resolution dependence on spectral frequency in human speech cortex electrocorticography. J. Neural Eng. *13*, 056013. https://doi.org/10.1088/1741-2560/13/5/056013.

67. Leszczyński, M., Barczak, A., Kajikawa, Y., Ulbert, I., Falchier, A.Y., Tal, I., Haegens, S., Melloni, L., Knight, R.T., and Schroeder, C.E. (2020). Dissociation of broadband high-frequency activity and neuronal firing in the neocortex. Sci. Adv. *6*, eabb0977. https://doi.org/10.1126/sciadv.abb0977.

68. Churchland, M.M., Yu, B.M., Cunningham, J.P., Sugrue, L.P., Cohen, M.R., Corrado, G.S., Newsome, W.T., Clark, A.M., Hosseini, P., Scott, B.B., et al. (2010). Stimulus onset quenches neural variability: a widespread cortical phenomenon. Nat. Neurosci. *13*, 369–378. https://doi.org/10.1038/nn.2501.

69. Arazi, A., Yeshurun, Y., and Dinstein, I. (2019). Neural Variability Is Quenched by Attention. J. Neurosci. *39*, 5975–5985. https://doi.org/10.1523/JNEUROSCI.0355-19.2019.

70. Arazi, A., Censor, N., and Dinstein, I. (2017). Neural Variability Quenching Predicts Individual Perceptual Abilities. J. Neurosci. *37*, 97–109. https://doi.org/10.1523/JNEUROSCI.1671-16.2016.

71. Mukamel, R., Gelbard, H., Arieli, A., Hasson, U., Fried, I., and Malach, R. (2005). Coupling Between Neuronal Firing, Field Potentials, and fMRI in Human Auditory Cortex. Science *309*, 951–954. https://doi.org/10.1126/science.1110913.

72. Huth, A.G., de Heer, W.A., Griffiths, T.L., Theunissen, F.E., and Gallant, J.L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. Nature *532*, 453–458. https://doi.org/10.1038/nature17637.

73. Nishimoto, S., Vu, A.T., Naselaris, T., Benjamini, Y., Yu, B., and Gallant, J.L. (2011). Reconstructing Visual Experiences from Brain Activity Evoked by Natural Movies. Curr. Biol. *21*, 1641–1646. https://doi.org/10.1016/j.cub.2011.08.031.

74. Broderick, M.P., Anderson, A.J., Di Liberto, G.M., Crosse, M.J., and Lalor, E.C. (2018). Electrophysiological Correlates of Semantic Dissimilarity Reflect the Comprehension of Natural, Narrative Speech. Curr. Biol. *28*, 803-809.e3. https://doi.org/10.1016/j.cub.2018.01.080.

75. Fair, D.A., Schlaggar, B.L., Cohen, A.L., Miezin, F.M., Dosenbach, N.U.F., Wenger, K.K., Fox, M.D., Snyder, A.Z., Raichle, M.E., and Petersen, S.E. (2007). A method for using blocked and event-related fMRI data to study “resting state” functional connectivity. NeuroImage *35*, 396–405. https://doi.org/10.1016/j.neuroimage.2006.11.051.

76. Rissman, J., Gazzaley, A., and D’Esposito, M. (2004). Measuring functional connectivity during distinct stages of a cognitive task. NeuroImage *23*, 752–763. https://doi.org/10.1016/j.neuroimage.2004.06.035.

77. Ryali, S., Supekar, K., Chen, T., and Menon, V. (2011). Multivariate dynamical systems models for estimating causal interactions in fMRI. NeuroImage *54*, 807–823. https://doi.org/10.1016/j.neuroimage.2010.09.052.