NeuroImage xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Extracting multidimensional stimulus-response correlations using hybrid encoding-decoding of neural activity

Jacek P. Dmochowski[a,*], Jason J. Ki[a], Paul DeGuzman[b], Paul Sajda[c], Lucas C. Parra[a]

[a] Department of Biomedical Engineering, City College of New York, New York, NY 10031, United States
[b] Neuromatters LLC, New York, NY 10038, United States
[c] Department of Biomedical Engineering, Columbia University, New York, NY, 10027, United States

ABSTRACT

In neuroscience, stimulus-response relationships have traditionally been analyzed using either encoding or decoding models. Here we propose a hybrid approach that decomposes neural activity into multiple components, each representing a portion of the stimulus. The technique is implemented via canonical correlation analysis (CCA) by temporally filtering the stimulus (encoding) and spatially filtering the neural responses (decoding) such that the resulting components are maximally correlated. In contrast to existing methods, this approach recovers multiple correlated stimulus-response pairs, and thus affords a richer, multidimensional analysis of neural representations. We first validated the technique's ability to recover multiple stimulus-driven components using electroencephalographic (EEG) data simulated with a finite element model of the head. We then applied the technique to real EEG responses to auditory and audiovisual narratives experienced identically across subjects, as well as uniquely experienced video game play. During narratives, both auditory and visual stimulus-response correlations (SRC) were modulated by attention and tracked inter-subject correlations. During video game play, SRC varied with game difficulty and the presence of a dual task. Interestingly, the strongest component extracted for visual and auditory features of film clips had nearly identical spatial distributions, suggesting that the predominant encephalographic response to naturalistic stimuli is supramodal. The diversity of these findings demonstrates the utility of measuring multidimensional SRC via hybrid encoding-decoding.

Introduction

Understanding the relationship between a sensory stimulus and the resulting neural response is a fundamental goal of neuroscience. Two distinct paradigms have shaped the pursuit of the neural code. The *encoding* approach attempts to explain neural responses from features of the stimulus, typically via linear filtering (Dayan and Abbott, 2001). Examples include receptive fields and spike-triggered averages in single-unit electrophysiology (Dayan and Abbott, 2001), the generalized linear model (GLM) in functional magnetic resonance imaging (fMRI) (Friston et al., 1994; Monti, 2011), spectrotemporal response functions (STRF) in electrocorticograms (Ding and Simon, 2012a), and temporal response functions in encephalographic recordings (Lalor et al., 2006; Lalor and Foxe, 2010; Liberto et al., 2015). In contrast to encoding, the *decoding* approach is to predict the stimulus by filtering over an array of neural responses. Decoding techniques have been shown to reconstruct experienced stimuli in a large number of findings spanning animal (Bialek et al., 1991; Warland et al., 1997; Stanley et al., 1999; Butts et al., 2007) and human investigations of both visual (Norman et al., 2006; Thirion et al., 2006; Miyawaki et al., 2008; Kay et al., 2008; Nishimoto et al., 2011; Horikawa et al., 2013) and auditory stimuli (Pasley et al., 2012; Luo and Poeppel, 2007; Mesgarani et al., 2014; Mesgarani and Chang, 2012; O'Sullivan et al., 2014).

The encoding and decoding approaches possess contrasting strengths and weaknesses: whereas encoding models operate on the stimulus and are thus easily interpretable (Naselaris et al., 2011), they generally predict the responses of individual data channels (i.e, neurons, voxels, or electrodes) and do not efficiently recover distributed neural representations. Decoding techniques filter neural activity over multiple channels and are therefore naturally suited to capturing distributed representations, but at the expense of models that are often difficult to interpret and prone to overfitting. Therefore, an approach that efficiently captures distributed neural representations and is readily interpretable in the stimulus space is lacking.

Here we propose a hybrid approach that combines the strengths of encoding and decoding. The technique integrates neural responses across space while filtering the stimulus in time, i.e. it "decodes" neural activity to recover an "encoded" version of the stimulus. By jointly

* Corresponding author.

http://dx.doi.org/10.1016/j.neuroimage.2017.05.037
Accepted 17 May 2017

learning decoding and encoding models, distributed neural representations are identified and explicitly linked to portions of the stimulus. In contrast to existing paradigms, this approach decomposes the neural representation of stimuli into multiple dimensions, with each dimension defined by a (spatial) response component and a (temporal) stimulus component.

To validate the ability of the proposed technique to recover multiple simultaneous stimulus-driven components, we first conducted a simulation study using data generated from a finite element model (FEM) of the head. The recovered components matched the ground-truth activations in both spatial topography and time course. We then evaluated the technique on recordings of neural activity in response to various naturalistic audiovisual stimuli and found multiple significant dimensions of stimulus-response correlation (SRC) for both auditory and visual features. These multiple dimensions were modulated by the attentional state of the observer. Interestingly, we found that independent visual and auditory features possessed a common response component, suggesting that the dominant EEG response to natural stimuli is supramodal.

Inter-subject correlations (ISC) of neural responses to natural stimuli have recently been shown to reflect a variety of behaviors (Hasson et al., 2008a, 2008b; Lahnakoski et al., 2014; Wang et al., 2015; Stephens et al., 2010; Dmochowski et al., 2012, 2014). We reasoned that stimulus responses would be similar across subjects and thus predicted that SRC would track ISC; indeed, SRC covaried with ISC for both auditory and visual features. In contrast to ISC and event-related potentials, however, the proposed method does not require repeated exposures to the stimulus and is thus applicable to the study of unique stimuli. We therefore studied neural activity during video game play and identified SRCs that reflected both the game difficulty and attentional state of the player. The variety of novel findings attests to the utility of the hybrid encoding-decoding approach.

Analytic methods

We develop the proposed technique by relating it to the two predominant approaches for the analysis of neural signals: predicting the neural response from the stimulus (encoding), and recovering the stimulus from the neural response (decoding).

Encoding: modeling the mapping from stimulus to response

Consider a stimulus whose time-varying features are encapsulated by signal $s(t)$. For an auditory stimulus, the values of $s(t)$ may represent the sound pressure envelope. For a visual stimulus, $s(t)$ may represent the luminance. The stimulus is presented to an observer, generating a neural response $r_i(t)$ in the ith data channel (for example, an electrode in a microelectrode or EEG array, or a voxel in fMRI). Encoding seeks to identify the mapping from $s(t)$ to $r_i(t)$. This is conventionally performed by filtering $s(t)$ to produce an estimated neural response $\hat{r}_i(t)$ for each channel:

$$\hat{r}_i(t) = h_i(t) * s(t). \tag{1}$$

For EEG, the filters $h_i(t)$ represent the evoked response for each electrode i (Lalor et al., 2006; Lalor and Foxe, 2010). Encoding filters $h_i(t)$ are generally found by maximizing the correlation between the observed neural response $r_i(t)$ and the estimated neural response $\hat{r}_i(t)$ (or a convolved version of $\hat{r}_i(t)$ in the case of generalized linear models used in fMRI (Friston et al., 1994)). Here the symbol * denotes temporal convolution, such that $h(t) * s(t) = \sum_\tau h(\tau) s(t-\tau)$, but it could also represent spatial convolution to model visual receptive fields (Dayan and Abbott, 2001). Note that this optimization problem is generally solved separately for each channel $i = 1 \ldots D$. Thus, the encoding approach does not leverage potentially distributed representations where the stimulus elicits correlated responses across multiple channels. In particular, if the response magnitude in a given channel is below statistical detection thresholds, it may be missed by an encoding approach.

Decoding: recovering the stimulus from the response

To remedy this, decoding techniques combine the neural responses of multiple channels and aim to reconstruct the stimulus (e.g. (Mesgarani and Chang, 2012; Friston et al., 1994)):

$$\hat{s}(t) = \sum_i w_i(t) \star r_i(t). \tag{2}$$

The decoding weights $w_i(t)$ perform both spatial *and* temporal filtering, and are found by maximizing the correlation between the observed stimulus $s(t)$ and the estimated stimulus $\hat{s}(t)$. Here the symbol $\star$ denotes temporal correlation, such that $w(t) \star r(t) = \sum_\tau w(\tau) r(t+\tau)$, captures the response after stimulus presentation at time t. Note that the responses of multiple channels are linearly combined to recover a single stimulus feature, and that the model parameters are now optimized jointly. However, one limitation of this technique is that it is difficult to directly interpret the decoding coefficients $w_i(t)$ (Haufe et al., 2014). Moreover, in cases with many channels and long temporal apertures, conventional decoding techniques are prone to over-fitting and thus require careful model regularization (Yamashita et al., 2008; Pereira et al., 2009; Pasley et al., 2012; Haxby et al., 2014).

It should be noted that both encoding and decoding techniques can be inverted (to become decoders or encoders) if the unconditional distributions of the stimulus or the response, respectively, are available (Naselaris et al., 2011; Nishimoto et al., 2011; Naselaris et al., 2009). While the statistics of the stimulus may be readily estimated, it may be more challenging to estimate the statistics of the response independently of the stimulus.

Hybrid encoding and decoding

Here we propose a combination of the encoding and decoding approaches by simultaneously filtering the stimulus in time and the neural responses in space:

$$\hat{u}(t) = h(t) * s(t), \tag{3}$$

$$\hat{v}(t) = \sum_i w_i r_i(t). \tag{4}$$

The encoder $h(t)$ and decoder w_i are found by maximizing the correlation between the encoded stimulus $\hat{u}(t)$ and the decoded response $\hat{v}(t)$. Note that the decoding is now purely spatial (accounting for distributed neural representations), while the encoding is purely temporal (capturing only the relevant portions of the stimulus). It is straightforward to expand both filtering operations to become spatio-temporal, at the cost of increased dimensionality. The proposed approach is depicted diagrammatically in Fig. 1.

We summarize the three approaches here so that the analogies can be more clearly identified:

$$r_i(t) \sim s(t) * h_i(t) \quad \text{(Encoding)} \tag{5}$$

$$\sum_i w_i(t) \star r_i(t) \sim s(t) \quad \text{(Decoding)} \tag{6}$$

$$\sum_i w_i r_i(t) \sim s(t) * h(t) \quad \text{(Hybrid)} \tag{7}$$

where, $u(t) \sim v(t)$ indicates that model parameters are selected to maximize the correlation between the signals $u(t)$ and $v(t)$:

$$\rho(u, v) = \frac{\sum_t u(t) v(t)}{\sqrt{\sum_t u^2(t) \sum_t v^2(t)}}, \tag{8}$$

and where zero-mean has been assumed for u and v. Model parameters that maximize correlation in (5) and (6) are unique and can be found

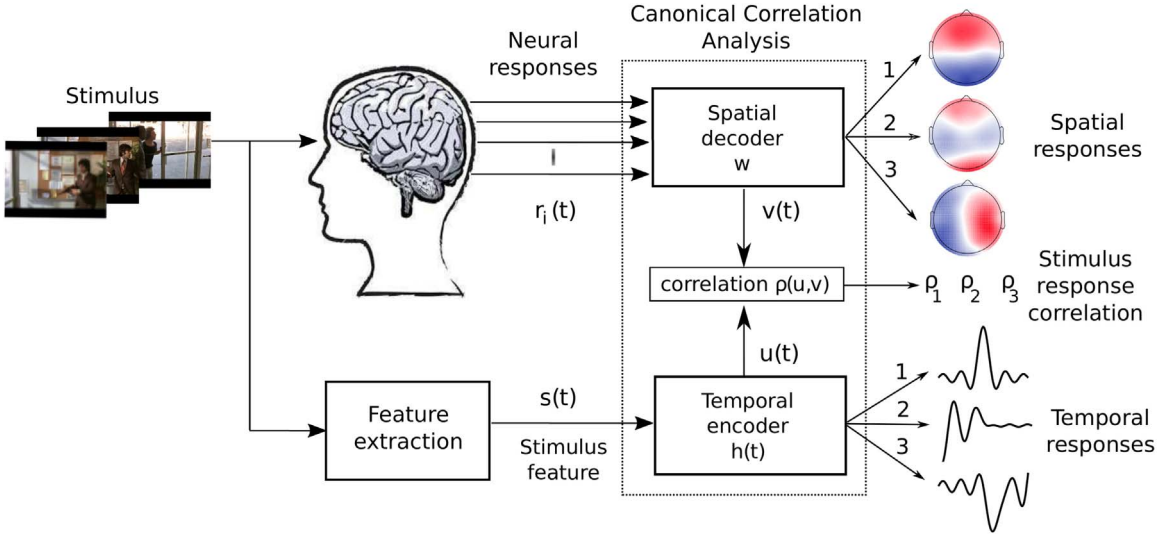

Fig. 1. Schematic view of the proposed technique. A stimulus impinges on the observer, generating a set of neural responses $r_i(t)$. The relevant features of the stimulus (for example, the time-varying luminance or sound envelope) are extracted, resulting in a time series $s(t)$. An optimization procedure, implemented here via canonical correlation analysis, then computes spatial filters w_k to apply to the neural responses and temporal filters $h_k(t)$ to apply to the stimulus features such that the resulting filter outputs are maximally correlated in time. The result is a set of multiple stimulus and response components whose activities track each other.

with conventional least-squares optimization, arguably leading to the popularity of the conventional encoding and decoding approaches (see *Methods*).

Optimizing the parameters of the hybrid technique (7) is performed by Canonical Correlation Analysis (CCA) (Hotelling, 1936), which provides *multiple* independent dimensions, or "components" of the correlation between the stimulus and the response. Specifically, each dimension $k = 1,\ldots,K$ is defined by an encoder $h_k(t)$ and decoder w_{ki}. Each encoding/decoding dimension k captures in $\hat{v}_k(t)$ a spatial component of the neural activity and in $\hat{u}_k(t)$ a temporal component of the stimulus. The SRC of dimension k is then defined according to:

$$\rho_k = \rho(\hat{u}_k, \hat{v}_k), \tag{9}$$

with $\rho_1 > \rho_2 > \ldots > \rho_K$, and zero cross-correlation across components, $\rho(\hat{u}_k(t), \hat{v}_l(t)) = 0$ for $k \neq l$ (this is approximately true also in the case of regularization, as demonstrated in the Results). Thus, different components capture genuinely different aspects of the stimulus-response relationship. Note that this multidimensional representation cannot be recovered with the either the encoding approach (5) or the decoding approach (6), as they inherently yield only one dimension of the SRC. Some have used principal components analysis (PCA) as a post-processing of encoding models to capture components of distributed representations (Huth et al., 2012, 2016). However, PCA enforces orthogonality on the weights of the spatial (or temporal when performing temporal component analysis) filters. In contrast, CCA only requires temporally uncorrelated filter *outputs*, and spatial filters w_i are not required to be orthogonal. We also note that CCA has been previously used to measure correlation between stimuli and MEG responses using purely temporal filters (Koskinen et al., 2013) without capturing distributed responses.

There are two conceptual differences between the hybrid approach (7) and conventional encoding or decoding (5)–(6). First, the hybrid approach captures distributed representations while providing readily interpretable stimulus response models. Second, the neural responses are separated into components k that are uncorrelated from each-other. In total, the neural responses to the stimulus are implicitly modeled as:

$$\hat{r}_i(t) = \sum_k a_{ki} h_k(t) * s(t). \tag{10}$$

Here the "temporal response" $h_k(t)$ represents the time course of neural activity evoked by the stimulus for component k. These temporal responses do not depend on space (electrode). The factor a_{ik}, which varies with electrode i, reflects how strong (and with which sign) the neural response is expressed across space. It thus captures the "spatial response". As with other decoding and source separation methods, it is easier to interpret the spatial responses a_{ik} than the spatial filter weights w_{ik} Haufe et al. (2014). Differences between filter weights at low-variance or noisy electrodes do not indicate genuine differences in neural responses to the stimulus. On the other hand, the spatial response captures the distribution of stimulus-related activity that is extracted by the spatial filter (see Appendix A, Eq. (16)).

Note that in the traditional encoding and decoding approaches (5)–(6), the number of free parameters is QD, with Q representing the length of the temporal aperture and D denoting the number of neural data channels. On the other hand, the hybrid method as formulated here leads to simpler models with $(Q + D)K$ free parameters. Typically, the total correlation is contained within a few components K (e.g. 5 or less) while D and Q often span hundreds of parameters. We thus expect the new technique to be less susceptible to overfitting the data.

Experimental methods

The proposed technique was evaluated using both simulated data as well as multiple real EEG data sets collected in various settings.

Simulation study

We conducted a simulation study of the proposed technique using a FEM of the segmented human head (Huang et al., 2016). The model included 10,000 EEG sources arranged along the grey matter surface and 230 electrodes placed on the scalp following an extended 10/5 scheme (Huang et al., 2016). The stimulus was extracted as the optical flow time series from a clip of the film "Dog Day Afternoon" (duration 325 s, 24 frames/s). This stimulus was encoded by two brain regions: the primary visual cortex (V1; coordinates in the head model −0.0078, −0.10,0.0028), whose temporal response to the stimulus was given by an impulse response shaped according to the Cauchy probability density function (PDF) with a peak at 167 ms and a scale parameter of 0.5. Visual area V4 (coordinates −0.026, −0.10, −0.0048) also represented the optical flow stimulus with an impulse response shaped according to a Cauchy PDF with a peak at 500 ms and a scale parameter of 1. Both temporal responses were normalized to unit L^2 norm, and are shown in Fig. 2A. A third brain region, modeled here as the inferior temporal cortex (ITC;

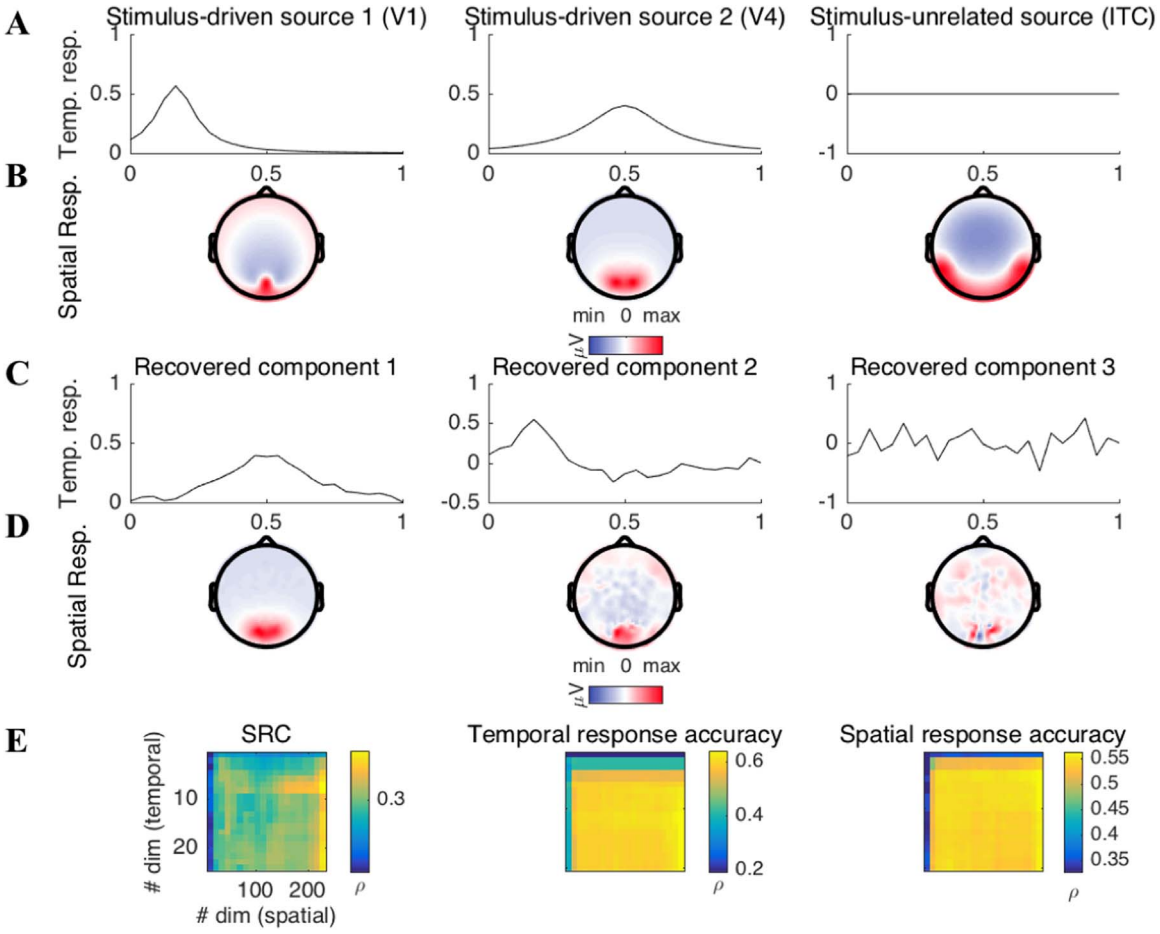

Fig. 2. Hybrid encoding-decoding recovers multiple dimensions of SRC. Simulated activations of the primary visual cortex and visual area V4 followed the optical flow of a film clip with temporal responses depicted in (**A**) and spatial responses depicted in (**B**). A source in the inferior temporal cortex (ITC) was uncorrelated with the stimulus. The hybrid technique recovered the temporal (**C**) and spatial (**D**) responses of the V4 source in component 1, and the V1 source in component 2, respectively, with correlations between the estimated and true responses being $r > 0.99$ (spatial) and $r > 0.68$ (temporal). (**E**) Regularization was performed here by truncating the eigenvalue spectrum of the stimulus (temporal) and response (spatial) covariance. Peak SRC was attained with a 9-dimensional stimulus covariance and a full (230) dimensional EEG covariance. The accuracy of the estimated spatial and temporal responses were relatively insensitive to regularization, potentially due to the presence of spatially white noise added to the EEG which effectively regularized the data.

coordinates −0.062, −0.035, −0.028), generated a stimulus-independent white noise waveform with a standard deviation of twice that of the stimulus-driven sources. Each source consisted of 40 vertices closest to the manually identified region coordinates. The three source activations were projected onto the scalp electrodes using the head model's lead field matrix and superimposed. Spatially white Gaussian noise was then added to the electrodes such that the signal-to-noise ratio (SNR) was 0.3 (−10 dB). The scalp projections (i.e., spatial responses) from the three modeled brain regions are shown in Fig. 2B. To evaluate the effect of regularization on the recovered sources and the SRC, we varied the strength of regularization (i.e., the number of principal components retained in the covariance matrices of the stimulus feature and EEG response; see *Regularization* in Appendix A) from 5% to 100% in 20 equally spaced intervals. The first 300 s of the stimulus and simulated EEG response were used to learn the model parameters, while the last 25 s served as the test set where performance was evaluated. We measured the SRC, the correlation between the recovered and actual spatial response, and the correlation between the recovered and actual spatial temporal response, all averaged over the first two components.

Subjects and stimuli

All participants provided written informed consent in accordance with the procedures approved by the Western Institutional Review Board (Puyallup, WA). 30 healthy human subjects (15 females, median age 23) participated in the main experiment, during which they freely viewed a clip from the film “Dog Day Afternoon” (duration 325 s, 24 frames/s; this clip was first analyzed in Honey et al. (2012) using fMRI) while their EEG was recorded. An additional 5 healthy human subjects (2 females, median age 21) were recruited for the video game study, where EEG was collected while subjects played the car-racing video game “SuperTuxKart” (variable duration, 60 frames/s). Subjects controlled the vehicle using their right hand, via keyboard directional arrows - left/right keys controlled steering, while up/down controlled acceleration. The video game study also included a “divided-attention” condition during which participants performed a concurrent rapid-serial-visual-presentation (RSVP) task (Gerson et al., 2006) to earn various “items” which gave players a temporary competitive advantage against the other racers. For this condition, a square black panel was superimposed on the screen with objects rapidly flashed (i.e., 5 Hz) inside the square. Subjects were instructed to attend (while maintaining eye gaze on their vehicle and race track) to a particular object of the RSVP display. Players could redeem their selected items by pressing the space bar with the left hand. Two levels of race difficulty were tested: “easy” and “hard”. The easy condition consisted of slow driving speed against 3 simulated race competitors, tuned to allow the subject ample opportunity to win the race. In the hard condition, the driving speed increased, as well as the number (i.e., 7) and performance of the simulated race competitors. This resulted in more obstacles, crashes, and aggressive driving. In both experiments, sound was delivered via Sony MDR-7506 headphones adjusted by each subject to a comfortable listening volume prior to the experiment.

Existing data

To evaluate the effect of attention (see Fig. 6), we reanalyzed the EEG data from (Ki et al., 2016). In that study, EEG was collected during passive viewing/listening of the following popular stimuli: “Bang! You're Dead” (N = 20, 8 females, median age 20, duration 372 s, 25 frames/s) (Hasson et al., 2008c), “The Good, the Bad, and the Ugly” (same subject pool, duration 388 s, 30 frames/s) (Hasson et al., 2004), and “Pie Man” (recorded on a separate N = 20 subjects, 7 females, median age 21, duration 360 s, 30 frames/s, (Lerner et al., 2011), although this audio narration only showed a fixation cross on the screen). There were two experimental conditions (each with N = 20): in the “attend” condition, subjects were instructed to normally attend to the stimuli. To emulate the inattentive state, in the “count” condition subjects were instructed to mentally count backwards in steps of 7 during viewing/listening.

EEG collection and pre-processing

Subjects were fitted with a 32-electrode cap placed on the scalp according to a modified 10/10 scheme for EEG, which was recorded with a BioSemi ActiveTwo system (BioSemi, Amsterdam, Netherlands) at a sampling frequency of 2048 Hz and 24 bits per sample. Four-channel electrooculogram (EOG) recordings were collected from electrodes below and adjacent to each eye. EEG pre-processing was performed automatically and offline in the MATLAB software (MathWorks, Natick, MA). The signals were high-pass filtered at 1 Hz, notch filtered at 60 Hz, and then downsampled to 256 Hz. To remove the contribution of eye movements from the EEG, the four EOG channels were linearly regressed out of the 32-channel EEG. Artifactual channels and data samples were identified and replaced with zeros when their respective power exceeded the mean power by 4 standard deviations. The EEG was further downsampled to the frame rate of the stimulus prior to analysis.

Stimulus feature extraction

All stimuli were loaded into the MATLAB software to extract the video frames and audio samples comprising the stimulus. The color video frames were converted to grayscale, resulting in intensity values $I_p(t)$ for pixel p at frame time t. The luminance at each frame was then computed as the mean intensity across pixels: $L(t) = \langle I_p(t) \rangle_p = \frac{1}{P} \sum_{p=1}^{P} I_p(t)$. Similarly, temporal contrast was derived as the mean temporal derivative of intensity changes, $\langle |\partial I_p(t)/\partial t| \rangle_p$ (i.e., unsigned frame to frame changes in intensity). Local contrast was computed following (Groen et al., 2013): $\langle |I_p(t) - H_p * I_p(t)| \rangle_p$, where * indicates a 2-dimensional spatial convolution, here with uniform point-spread function H_p with a 30×30 region-of-support. Optical flow was estimated from the frame sequence using the Horn-Schunk method (Horn and Schunck, 1981) via the MATLAB Computer Vision System Toolbox. The sound envelope was computed as the squared magnitude of the Hilbert transform of the sound pressure amplitude, and then downsampled to the frame rate of the accompanying video. Prior to processing, all features were high-pass filtered at 1 Hz (to match the neural response and remove slow drifts) and z-scored.

Cross-validation and statistical significance

When learning the optimal hybrid model parameters, we performed leave-one-out cross-validation along the subject dimension. In particular, we held out one subject at a time, learning the encoding ($h_k(t)$) and decoding filters (w_{ki}) on the data pooled from the remaining subjects. The resulting model parameters were then applied to the held-out subject to measure the SRC on “unseen” data.

Cross-validation was not performed for the SRC-ISC comparison (Fig 5) due to the fact that the goal of the analysis was to measure the correspondence between SRC and ISC: any spurious model fits would not be expected to produce SRCs that covary with the ISC.

In order to determine statistical significance of the SRC at each learned component pair ($\rho_k > 0$), we formed N = 1000 surrogate data records in which the phase spectrum of the EEG was randomized following (Theiler et al., 1992). This procedure preserved the auto-correlation structure of the EEG while disrupting the temporal relationship between the stimulus and neural activity. SRC computed with the permuted data records defined the null distribution from which p-values were estimated.

SRC-ISC comparison

When comparing SRC with ISC (Fig. 5), we learned the encoding and decoding filters separately for each individual subject (i.e., no pooling of subjects' EEG was conducted for this analysis). This was performed in order to ensure that the computation of SRC-maximizing filters would not be biased towards picking up patterns of activity that were common across subjects, thus confounding the SRC-ISC relationship. For each subject, their SRC-maximizing spatial filter was applied to their EEG record, and the resulting EEG components were then correlated with their corresponding optimally filtered stimulus feature. The correlation was computed in a time-resolved fashion using windows of 5-second length and 80% overlap across successive windows. At each window, the SRC was uniformly summed across the first three components in order to reduce the dimensionality of the comparison. To compute ISCs of the neural responses, we followed the procedure of (Dmochowski et al., 2012). The subject-independent spatial filters that maximized ISC across the subject pool were learned and then applied onto each subject's data before computing pairwise ISCs and then summing across all $N \times (N-1)/2 = 435$ subject pairs. As with the SRC, the ISC was computed across 5 s windows and uniformly summed across the first three components.

To test for statistical significance of the correlation between ISC and SRC, we performed a permutation test where the phase-spectrum of the ISC time series was randomized (Theiler et al., 1992). By preserving the magnitude spectrum, this procedure maintains the autocorrelation of the ISC signal. The permuted ISC was then correlated with the SRC, and the procedure was repeated 100,000 times. The p-value was then estimated from the proportion of iterations in which the mock correlation exceeded the true one.

Results

Hybrid encoding-decoding recovers multidimensional SRC

To verify that the proposed hybrid technique can recover multi-dimensional SRC, we conducted a simulation study in which two visual brain regions represented the optical flow of a film stimulus with differing delays. The temporal responses of the primary visual cortex (V1) and visual area V4 are shown in Fig 2A, along with the response of the inferior temporal cortex (ITC) whose activation was uncorrelated with the stimulus waveform. The spatial responses of these activations are shown in Fig 2B and represent the projections of the activated brain regions on the scalp. We expected that the topographies of the stimulus-driven sources would be recovered by the proposed technique, but not the topography of the stimulus-unrelated source.

The first component recovered by the hybrid technique closely matched the V4 activation in both temporal response (Fig 2C, ρ = 0.99, N = 25 with no regularization) and spatial response (Fig 2D, ρ = 0.99, N=230 with no regularization). The second recovered component matched the V1 activation (temporal response: ρ = 0.93, spatial response: ρ = 0.68). As expected, the spatial response of the stimulus-independent source was not recovered. To evaluate the effect of regularization on the recovered sources and SRCs, we repeated the analysis at different levels of regularization, measured here as the

number of principal component dimensions retained in the covariance matrices of the stimulus feature and EEG response (see *Regularization* in Appendix A for details). For each regularization level, we quantified the accuracy of the recovered spatial responses by computing the correlation with the actual projections from the cortical sources to the scalp. Similarly, we measured the correlation between the true temporal responses and the recovered ones. In each case, we averaged the correlations among the first two components. The SRC peaked when retaining 9 (of a possible 25) temporal stimulus components and all 230 spatial EEG dimensions (Fig 2E). The accuracies of the temporal and spatial responses were highest for a stimulus dimensionality of 14 and 8, respectively, with regularization of the EEG failing to improve the accuracy of the recovered components. This could be due to the fact that spatially white noise was added to the simulated EEG, thus effectively regularizing the covariance.

Note that there is an inherent sign ambiguity in the recovered spatial and temporal responses: each can be negated without altering the value of the resulting SRC. Therefore, in Fig 2 (e.g. component 2) and throughout, we have in some cases inverted the polarity of the recovered topographies and associated time courses to match the ground-truth or a component from another condition (e.g. feature).

In the following, we demonstrate the utility of hybrid encoding-decoding by applying the technique to real EEG responses evoked by naturalistic audiovisual stimuli.

Dynamic visual features elicit strong multidimensional SRC

We first sought to determine which features of naturalistic stimuli evoke the strongest SRC. For a popular film clip during which we recorded the evoked scalp potentials of $N = 30$ viewers, we extracted a set of visual and auditory features including optical flow, visual temporal contrast, sound amplitude envelope, luminance, and spatial contrast (see *Methods*). Applying these derived features and neural responses to the hybrid encoding-decoding scheme led to a range of SRC values (Fig. 3A). We found statistically significant correlations along multiple dimensions (i.e., component pairs) for four of the five features (all $p < 0.05$, computed using phase-randomized surrogate data). SRCs of different component pairs are shown cumulatively, as each pair captures a different dimension in the data with uncorrelated activity. The strongest SRCs were exhibited by temporal contrast and optical flow, exceeding the correlations with sound envelope (paired t-test for sound envelope with temporal contrast, $t(29) = 5.7$, $p = 4 \times 10^{-6}$, and with optical flow $t(29) = 4.7$, $p = 5 \times 10^{-5}$, both $N = 20$).

By construction, the response components recovered by CCA are temporally uncorrelated with one another. However, when regularizing covariance as was performed here (see *Methods*), the response components may exhibit some level of correlation. Thus, to confirm that regularization did not introduce "cross-talk", we also measured the correlation between mismatched stimulus and response components (e.g. the correlation between stimulus component 1 and response component 2). These correlations were found to be very low (mean ± standard deviation across all features and component pairs: 0.0005 ± 0.003). Comparing this to the SRC measured within matched component pairs (as high as 0.1 for temporal contrast), it is clear that the multiple response components detected by the hybrid technique were distinct.

Similar spatial responses to auditory and visual stimuli

The hybrid technique correlates spatial *response components* with temporal *stimulus* components. Response components are characterized by a topography, termed a *spatial response*, that conveys the spatial distribution of the decoded neural activity. Stimulus components are extracted by a temporal filter that conveys a *temporal response*. Together, the spatial and temporal responses fully convey the mapping between original stimulus and evoked neural response (see *Methods* for details). Interestingly, the spatial responses of the first component were similar for all features (Fig. 3B).

While the finding of congruent spatial responses was expected for the two dynamic visual features which were strongly correlated (correlation between optical flow and temporal contrast: $r = 0.96$), we did not expect to find such similar topographies for weakly correlated auditory and visual features (correlation between sound envelope and temporal contrast: $r = 0.067$). To rule out that the similarity of the auditory and visual spatial responses was due to this small correlation, we subtracted from the temporal contrast the fraction that was explained by the sound envelope, and vice versa. In doing so we generated orthogonal time series for temporal contrast and sound envelope. The spatial responses of the first component for these uncorrelated visual and auditory features still had nearly identical distributions on the scalp (Fig. 4, left; correlation between spatial responses of sound envelope and temporal contrast: $r = 0.99$). The associated temporal responses (filters) showed opposing polarities of

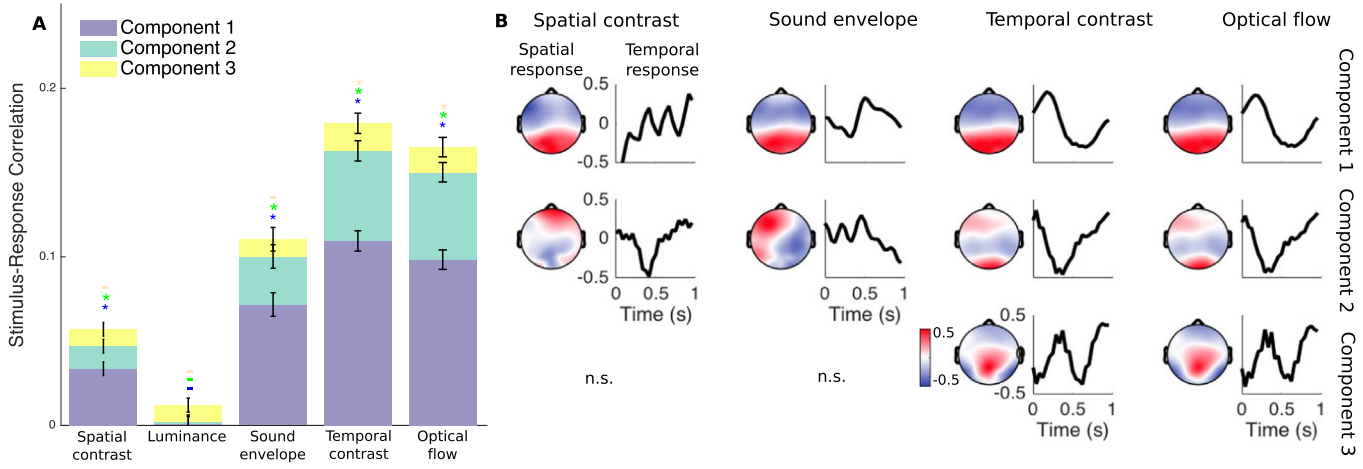

Fig. 3. Neural responses to a film clip track dynamic visual features. **A** SRC as computed for various auditory and visual features of a clip from "Dog Day Afternoon". Significant correlations were detected along multiple dimensions for 4 of the 5 features considered ($p < 0.05$, permutation test using surrogate data, and indicated by '*' with color indicating the component tested). Correlations with temporal contrast and optical flow, features that differentiated pixel values across frames, exceeded those with the sound envelope ($p < 0.04$, paired t-test, N=30). Error bars denote the standard errors of the mean (SEM) across subjects. **B** Spatial responses convey the topography of neural activity that best expressed the stimulus features. Temporal responses reflect the delays between stimulus and neural response. While the topographies of the first response components (left panels in top row) were congruent for all features, the temporal responses varied with the particular feature used.

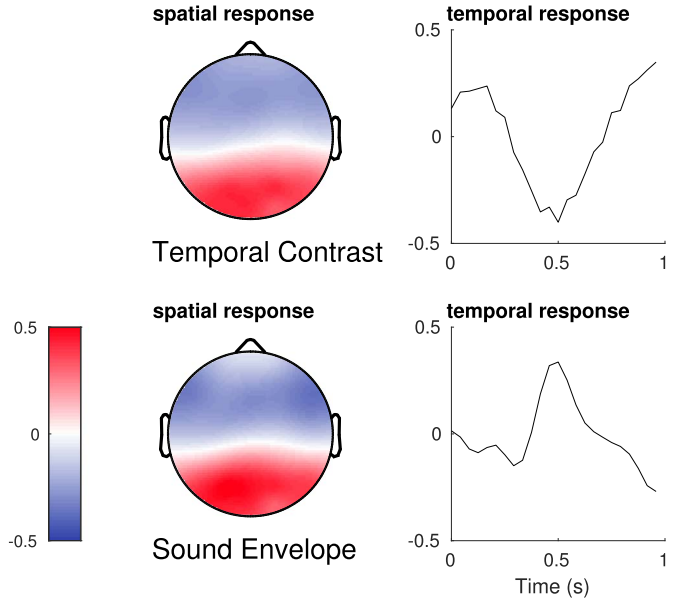

Fig. 4. Visual and auditory features evoke similar spatial responses. The spatial responses of the first component for visual temporal contrast and sound envelope, which depict where on the scalp these stimuli were expressed, were highly similar (left column, r = 0.99), even though the features were decorrelated (r=0). However, the associated temporal responses were inversely related (right column, r = −0.83), suggesting that the evoked responses to visual and auditory stimuli drove the EEG with opposing polarity. All values have arbitrary units as SRC is independent of scale.

the responses to visual and auditory features (Fig. 4, right). We checked whether these temporal filters introduced a correlation between the two features, and found only a weak negative correlation of $r = -0.14$ between filtered visual and auditory features. We also investigated whether the regularization added to the computation of the filter weights (see *Regularization* in Appendix A) may have led to the similarity of the spatial responses. After repeating the analysis but without regularization, we once again found that the spatial responses were highly similar for auditory and visual features (r = 0.97). Therefore, one possible interpretation of similar spatial responses to visual and auditory stimuli is that the dominant EEG response to natural stimuli is supramodal (however, see *Discussion*).

SRC tracks inter-subject correlation (ISC)

A number of reports have shown that dynamic natural stimuli elicit similar responses across subjects in fMRI, EEG and MEG (Hasson et al., 2008a; Dmochowski et al., 2014; Lankinen et al., 2014). For responses to be reproducible across subjects, the responses should also be reliably evoked by the stimulus within subjects. Therefore, we hypothesized that there would be a correspondence between how strongly the stimulus drove individual neural responses and how similar the responses were across subjects. To test this, we computed a *time-resolved* measure of the SRC (by summing across the first three component pairs of the hybrid technique) for the temporal contrast and sound envelope of the same film clip. Similarly, we also measured the time-resolved ISC (by summing across the three components maximizing correlation across subjects) experienced during the same stimulus (see *Methods*). In line with our hypothesis, a significant portion of the variability in the ISC time series could be explained from both visual and auditory SRC (Fig 5; temporal contrast: r = 0.59, $p = 3 \times 10^{-31}$, N = 321; sound envelope: r = 0.34, $p = 3 \times 10^{-10}$, N = 321). While these correlations are not large, it is still remarkable that the responses to simple unimodal stimulus features can explain a significant portion of the ISC variability.

Attentional modulation of SRC

Given the long-standing evidence showing that attention modulates evoked responses (e.g. (Picton and Hillyard, 1974)), we hypothesized that the SRC would decrease when the level of attention directed to the stimulus was reduced. To test this, we reanalyzed previous data recorded in two attentional conditions (Ki et al., 2016): subjects either naturally attended to the stimulus or performed a counting task, intended to distract viewers from the stimulus. Two film clips (“Bang! You're Dead” and “The Good, The Bad, and the Ugly”) and one narrated audiobook (“Pie Man”) were considered for this analysis. All subjects were presented the same stimuli under both conditions (i.e., we took repeated measures). We investigated whether SRC was modulated by attention (attend vs count) and if this was specific to particular stimuli or component pairs. We first analyzed the SRC using the sound envelope. A three-way, repeated-measures ANOVA with component, attention and stimulus as factors identified main effects of attention ($F(1) = 7.48$, p = 0.008) and stimulus ($F(1) = 29.17$, $p = 1.82 \times 10^{-9}$) and an interaction between component and stimulus ($F(2) = 14.15$, $p = 1.02 \times 10^{-5}$). Follow-up pairwise comparisons showed that the reduction in SRC for the “count” condition was driven by the two stimuli containing speech (Fig. 6A). In contrast, the “The Good, the Bad, and the Ugly” had minimal speech content and elicited weak SRC that was not modulated by attention. This suggests that the effect of attention on auditory EEG responses may be specific to speech. For the two audiovisual clips, we measured SRC using the optical flow and again performed a three-way, repeated-measures ANOVA with attention, stimulus and component as factors. There was again a strong main effect of attention ($F(1) = 34.6$, $p = 10^{-5}$), with reduced SRC in the “count” condition (Fig 6A). An interaction between attention and stimulus ($F(1) = 12.2$, p = 0.002) was also found. Specifically, SRC during the suspenseful “Bang! You're Dead” was more robustly modulated by attention, consistent with results reported in Ki et al. (2016).

SRC during interactive stimuli

To demonstrate that the proposed technique can capture SRC elicited by uniquely experienced stimuli, we recorded scalp potentials from N = 5 subjects while playing a car-racing video game (Fig 7A). The ongoing feedback between player and game meant that every race was perceptually unique. After reconstructing the optical flow of the video game display during each race, the stimulus time series and neural responses were used to measure SRC with the hybrid technique.

We found 5 statistically significant components whose spatial and temporal responses are illustrated in Fig 7B. Note that the spatial response of the first component was focused over parietal electrodes, and the associated temporal response had an early peak (i.e., 150 ms, Fig 7B). This is in contrast to the first component observed during film viewing (see Fig 3B). A component with a similar spatial but not temporal response as this “film” component was observed more weakly during video game play (Fig 7B, component 4). The discrepancy between film and video game components may indicate that when actively engaged with natural stimuli, distinct neural circuits are recruited. An alternative interpretation is that somatosensory and motor activity correlated with optical flow as players controlled speed and direction with right-hand key-presses. In this case, however, one would have expected a more central spatial response.

SRC is modulated by difficulty and presence of dual-task

The video game consisted of a car race with obstacles and competing drivers. Players experienced two levels of race difficulty by varying the number and skill of competing drivers. The game also included a divided-attention condition which required players to simultaneously attend to the top center of the screen (Fig. 7A), where items were presented for selection (Gerson et al., 2006). We predicted that players would devote

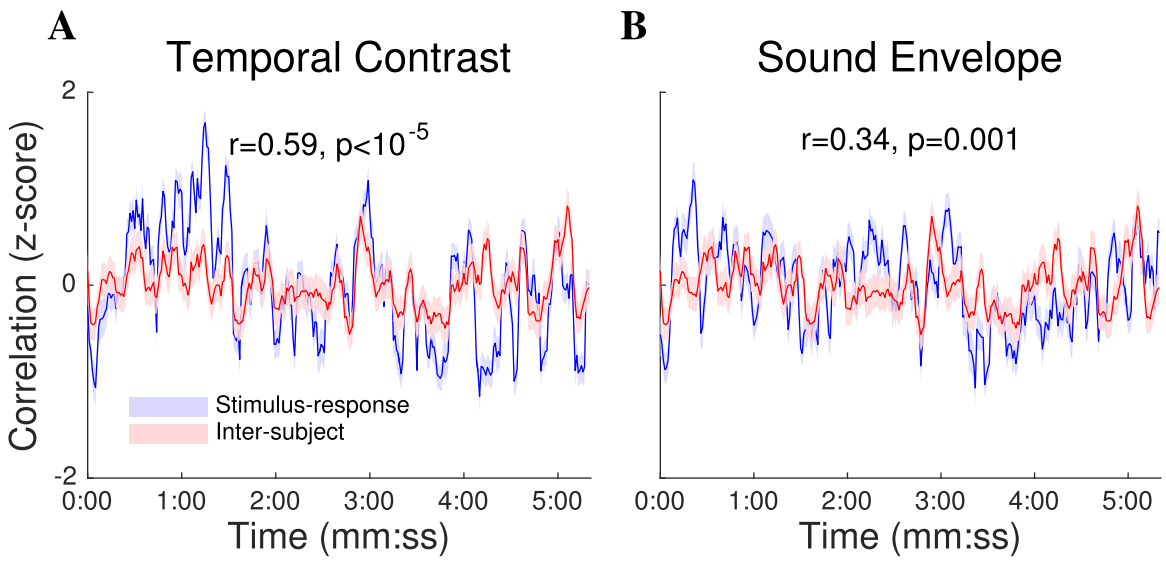

Fig. 5. SRC tracks the inter-subject correlation (ISC) of neural responses to naturalistic stimuli. **A** The time course of the SRC, as computed on the temporal contrast of a film clip, explains 34% of the variability in the ISC time course ($r = 0.59$, $p < 10^{-5}$, $N = 321$, permutation test). This suggests that the exogenous drive provided by the common stimulus underlies the reliability of neural responses across subjects as measured by the ISC. **B** Same as (A) but now for the envelope of the film's soundtrack. The SRC accounts for approximately 12% of the variability in the ISC ($r = 0.34$, $p = 0.001$, $N = 321$, permutation test). Shading indicates the SEM across subjects (for SRC) or subject pairs (for ISC).

more resources to the driving task during difficult races, resulting in higher SRC. In contrast, during periods of divided attention to a secondary task, SRC would be reduced. Repeated measures ANOVA with difficulty, attention, and component as factors revealed main effects of difficulty ($F(1) = 4.58$, $p = 0.035$), attention ($F(1) = 3.97$, $p = 0.050$), and component ($F(4) = 29.15$, $p = 0$ to numerical precision), with correlations increasing during difficult races and decreasing during the divided attention task, as predicted. It is important to note that the stimuli differed with game difficulty, and thus one cannot rule out that the effects are a result of varying stimuli and not of varying neural responses.

Discussion

Here we have developed a hybrid technique for learning the mapping between a dynamic stimulus and the corresponding neural response. By simultaneously encoding the stimulus and decoding the neural response, the proposed approach recovers multiple dimensions of SRC via the canonical correlation analysis formalism. We employed the technique to show that the brain's dominant response to visual and auditory stimuli had a common spatial response, even after removing all correlation between the film's soundtrack and visual features. Moreover, the dimensions of the SRC were modulated by the viewer's attentional state. The SRC was also shown to track the ISC of neural responses that have recently been employed to decode a variety of cognitive states. In contrast to the ISC, however, the multidimensional SRC does not require multiple subjects to experience the same stimulus. The technique is thus applicable to the study of interactive stimuli, as was demonstrated here for video game play, where both attentional and task demands were shown to modulate the SRC.

EEG tracks dynamic visual features. While there have been multiple reports of EEG responses tracking auditory features, in particular the envelope (Lalor and Foxe, 2010; Ding and Simon, 2012a, 2012b; Power et al., 2012; Golumbic et al., 2013; Liberto et al., 2015), relatively little comparable findings exist for visual stimuli (see Groen et al. (2013) for an exception). The hybrid technique developed here demonstrated that dynamic visual features were correlated with encephalographic responses to a level comparable if not stronger than the well-studied auditory envelope.

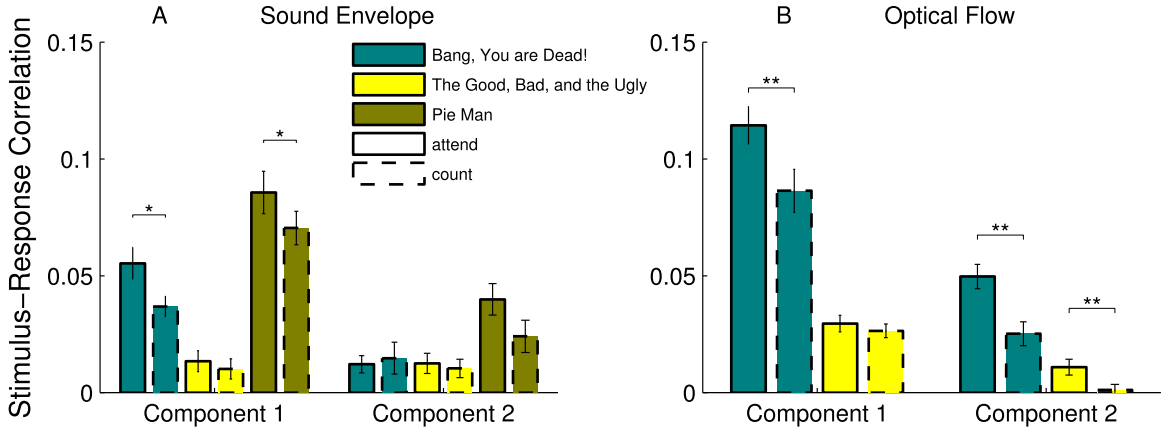

Fig. 6. Multidimensional SRC is modulated by the level of attention directed to the stimulus. Subjects viewed film clips or listened to an audiobook while either naturally attending to the stimulus ('attend') or performing a counting task ('count'). **A** SRC as measured for the sound envelope of two audiovisual and one auditory stimulus. Repeated measures ANOVA showed a significant effect of attention ($F(1) = 7.48$, $p = 0.008$). Follow-up comparisons indicated that SRC was modulated by attention in the first component but only for the stimuli that contained speech (t-test, BYD: p=0.021 and PM: $p = 0.044$, both $N = 20$). **B** Repeated measures ANOVA on the SRC with optical flow also found a main effect of attention ($F(1) = 34.6$, $p = 10^{-5}$). Follow-up comparisons indicated that this effect was robust in the first component for BYD (t-test: p=0.002, $N = 20$), and in the second component for both audiovisual stimuli (t-test, BYD: $p = 0.0011$ and GBU: $p = 0.0036$, both $N = 20$). Error bars denote SEM across subjects.

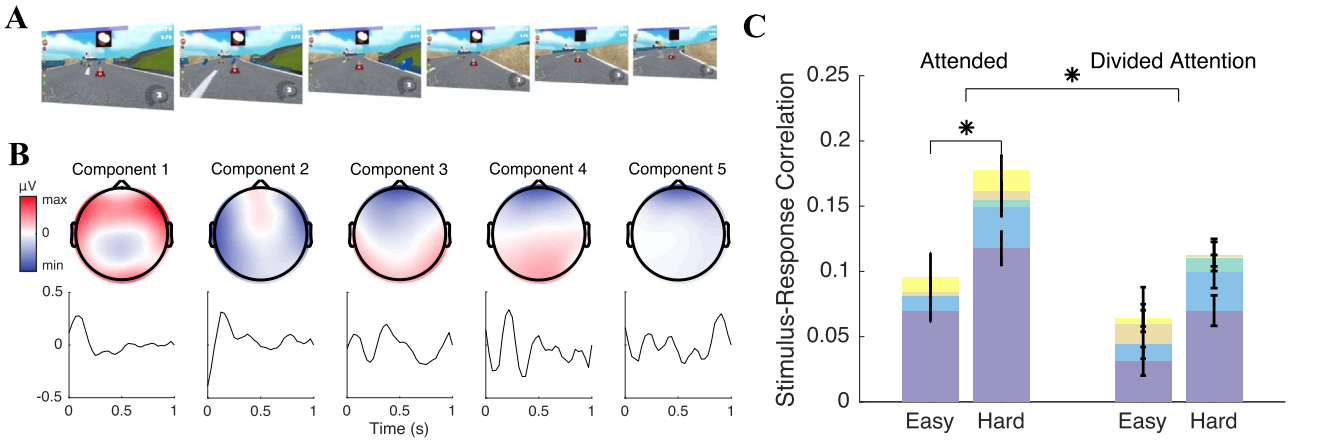

Fig. 7. SRC with optical flow measured during video game play is modulated by game difficulty and attention to a secondary task. **A** Neural activity was recorded while subjects played a car-racing video game under two difficulty levels and with some of the races containing an additional cognitive task. **B** The hybrid technique resolved five significant components, with the strongest component showing a parietal spatial response and an early temporal response (150 ms). **C** Repeated measures ANOVA showed main effects of difficulty ($F(1) = 4.58$, $p = 0.035$, asterisk displayed over bars in the 'attend' condition), attention ($F(1) = 3.97$, $p = 0.05$), and component ($F(4) = 29.15$, $p = 0$ to numerical precision). SRC was increased during difficult runs, while *decreasing* when the player's attention was divided between the game and a secondary task. Error bars denote the SEM across subjects.

Supramodal component. The supramodal component identified here bears some resemblance to the component that was previously found to maximize the ISC of EEG responses to video (Dmochowski et al., 2012, 2014) and auditory narratives (Ki et al., 2016; Cohen et al., 2016). In these earlier studies, the supramodality was obscured as the ISC approach is blind to which features of the stimulus drive reliable responses. Here we demonstrated that both auditory and visual features correlated with this component. It is thus possible that this component is selective to integrated audiovisual activity, as has been observed in temporal cortex during presentation of speech (Callan et al., 2001) as well as individual letters (Raij et al., 2000). Alternatively, the activity may be related to attentional networks that are entrained by the stimulus regardless of modality (Lakatos et al., 2009; Walz et al., 2013). It is also possible that the similarity of the spatial responses resulted from the correlation between auditory and visual features *after* temporal filtering (encoding). While this is unlikely as the correlation between filtered features was found here to be small, additional experiments are required to rule out this possible confound.

Distributed stimulus representations. A basic premise of the proposed approach is that stimulus features are represented by distributed rather than local neural responses. This is particularly true for EEG where the activity from a localized neural population can be detected at multiple electrodes. The encoding approach, conventionally used for analyzing spiking activity (Dayan and Abbott, 2001), fMRI (Friston et al., 1994) and recently also EEG (Lalor et al., 2006), models neural responses at individual channels (electrodes, voxels), and does not directly leverage such distributed activity. Decoding approaches, in contrast, can combine responses that are distributed and appear only weakly in individual channels (Kamitani and Tong, 2005; Norman et al., 2006). While encoding models are sometimes reversed to provide decoding (Nishimoto et al., 2011), such an approach often ignores the correlated nature of neural responses (Eyherabide and Samengo, 2013). The hybrid encoding-decoding technique captures distributed representations as components of the neural response. These components are linked to temporal components of the stimulus, and are thus readily interpretable.

Separating multiple dimensions of SRC. An important aspect of the proposed technique is its ability to extract multiple, independent dimensions of SRC. These multiple dimensions allow one to more finely probe the effects of experimental manipulations. For example, consider the effect of attention on SRC during film viewing (Fig. 6). For the sound envelope, only the correlation of the first component was modulated by altering attentional state. Conversely, for optical flow, the modulation was strongest in the space of the second component. Thus, the multidimensional nature of the proposed approach allows one to identify the neural circuits driven by experimental variables. The challenge of any decomposition technique, including principal, independent or correlated component analysis (Parra et al., 2005; de Cheveigné and Parra, 2014; Dmochowski et al., 2012), is to identify the functional significance of the extracted components. One approach is to manipulate the stimulus, task, or cognitive state of the subjects (as we have done here), and determine how the different components respond to such manipulations. Additionally, one can interpret the spatial response by comparing it against known functional anatomy (in the case of EEG/MEG, the spatial response can be projected onto cortex via an appropriate inverse model), or by comparing the delays in the temporal response with what is known about neural representation.

Relation to Inter-Subject Correlations. The SRC was shown to temporally covary with the ISC of neural responses to the stimulus. This is consistent with the idea that the common stimulus "synchronizes" the neural responses of multiple viewers. In principle, the ISC can be driven by any property of the stimulus, including high-level semantic features. Here we found that a significant fraction of the ISC fluctuation was explained by relatively simple unimodal features such as temporal contrast ($r = 0.59$) and sound envelope ($r = 0.34$). This result is significant because unlike the ISC approach, SRC may be measured with only one subject and one exposure to a stimulus. There are several recent examples of the ISC reflecting behavioral outcomes. For instance, ISC predicts subsequent memory of the stimulus (Hasson et al., 2008a), is indicative of viewer engagement (Dmochowski et al., 2012, 2014), correlates with the effectiveness of communication between individuals (Stephens et al., 2010; Silbert et al., 2014), and reveals the time scale of information integration for narratives (Hasson et al., 2008b). With the proposed technique, many of these studies can now potentially be conducted in the context of interactive stimuli. This is particularly useful with video game play, which is adaptive to user behavior and thus results in a different stimulus for every rendition of the game. Whether or not the SRC is also predictive of complex behaviors remains an empirical question. Our preliminary findings showing that attentional and task-demands modulate the SRC suggest that this is indeed the case.

Attentional modulation of evoked activity. Manipulating attentional state has been previously shown to strongly affect the ISC of the encephalogram (Ki et al., 2016), and preliminary evidence (Poulsen et al., 2017) suggests that this may partly result from varying evoked response magnitude, which is known to be affected by attentional state. We thus reanalyzed the data from (Ki et al., 2016) and found that both auditory and visual SRC were indeed reduced when directing attention

away from the stimulus (Fig. 6). Modulation of SRC with attentional tasks has been previously demonstrated for the cocktail party problem (i.e., attend to the voice of speaker A vs speaker B when both are simultaneously speaking). Both decoding (Mesgarani and Chang, 2012; O'Sullivan et al., 2015; Golumbic et al., 2013) and encoding approaches (Ding and Simon, 2012a) have been used in this context. It is interesting that the attentional modulation of the SRC with sound envelope was found here only for stimuli that contained speech, indicating that the modulation was not due to generic sound-evoked responses. We also found a robust modulation of SRC for the visual feature, but this too depended on the specific stimulus. Once again, this suggests that attentional modulation may not be a general property of evoked responses, and may explain the mixed results found in studies of task-related visual attention (O'Connell et al., 2009; Saupe et al., 2009).

The video-game experiment was designed to test the hypothesis that a more challenging game would also be more engaging. This would presumably increase the attentional focus of the player on the stimulus, resulting in an increase of SRC. In contrast, the distracting secondary task would reduce attention from the primary visual stimulus and thus reduce SRC. The modulations of SRC observed with task difficulty and presence of the dual task (Fig. 7) were consistent with this hypothesis. However, we cannot rule out that the effects on SRC were due to variability in the stimulus itself, which also changed along with the manipulations to attention and difficulty.

Interpreting video game activity. We found that the neural response to optical flow differed both spatially and temporally depending on whether the subject was passively observing or actively engaged with the stimulus. During active play of a video game, a response component with a parietal topography and an early time course emerged. This was in contrast to the slower supramodal component that was found to best correlate with optical flow during passive film viewing. While it is tempting to speculate that this result is evidence of mode-dependent visual processing, we cannot rule out alternative explanations that involve the effects of motor actions on the EEG: during video game play, subjects continually pressed keyboard buttons to control the game. Even though the observed spatial response is not consistent with a motor topography, there is evidence that button presses alter the task-evoked topographies of oddball paradigms (Salisbury et al., 2001). Further manipulations that control for the effects of key presses are needed to pin down the source of the response component recovered during video game play.

Owing to the similarities between film and video game stimuli, one may have expected to find the supramodal "film" component in the video game data. Indeed, we found that a spatially similar component emerged more weakly in the video game analysis (i.e., component 4). Note that while CCA does not impose orthogonality between spatial filters, it does require the activity of the various components to be uncorrelated. This decorrelation constraint complicates the comparison of components across different experiments (i.e., it is not straightforward to relate component 4 in the video game analysis to the supramodal component in the film experiments).

Comparing hybrid approach to encoding or decoding alone. One may naturally be tempted to compare the correlations achieved by the three general approaches (encoding, decoding, hybrid) on common data sets to determine which method works "best". However, these three approaches operate on different spaces: encoding correlates neural responses, decoding correlates stimuli, while the hybrid approach correlates filtered stimuli with filtered responses. As stimulus features and brain signals generally possess distinct noise characteristics, high correlation values do not necessarily indicate that one approach captures more of the relationship between stimulus and response than the others. For example, the hybrid approach may achieve a high correlation between stimulus and response components that contribute little to the variance in the *overall* stimulus and response. Therefore, a direct comparison of the correlations achieved by the three approaches would be difficult to interpret. Moreover, the choice of which approach to adopt will likely be guided by the particular application being considered.

Previous uses of CCA in neuroimaging. CCA has been used extensively to relate neural activity between several subjects or between different imaging modalities, but there are only isolated efforts aiming to capture stimulus-response relationships. Fujiwara et al. (2013) uses CCA to extracts patterns of activity in a local neighborhood of an fMRI voxel to correspond to linear combination of pixel intensities in a visual stimulus. This is very similar in spirit to the present work and captures responses that are distributed in space. However, their approach does not capture temporal responses as we have done here, nor can their formulation be readily expressed as a multidimensional, spatio-temporal encoding model – equations (10) or (16). Other related multivariate methods have been used to capture a distributed representation in the fMRI while at the same time linearly combining stimulus features (e.g. (Worsley et al., 1997; Friman et al., 2001; Nandy and Cordes, 2003; Kriegeskorte et al., 2006), summarized in Friston et al. (2008)), but they suffer from the same limitations, namely, no temporally delayed response is captured and it is not clear how to express the resulting model as a multidimensional encoding model.

Extensions to non-linear architectures and microelectrode arrays. Our ability to uncover the principles of sensory representation goes hand in hand with the ability to explain variance in the neural response. Here the relationship between stimulus and response was constrained to multiple linear mappings. It is expected that the incorporation of more sophisticated architectures that capture non-linear mappings will increase the magnitude of observed SRC. For example, deep neural networks that can synthesize complex functions and account for higher-order correlations may be implemented in a regression. A deep-learning extension of classical CCA has recently been formulated (Wang et al., 2015; Andrew et al., 2013). Kernel methods that exhibit robustness to overfitting may also prove useful (Akaho, 2001; Felix Bießmann et al., 2010). Finally, we note that the formalism presented here is equally applicable to other types of neural data including magnetoencephalography (MEG), fMRI, and multi-unit activity.

Appendix A

Implementing hybrid encoding-decoding

To implement the required optimization problems (Eq. (5)–(7)), we now formulate those expressions in matrix-vector notation. Let the stimulus be represented as a L-length row vector $\mathbf{s}$, where L is the duration of the stimulus and the neural response as a $D{\times}L$ matrix $\mathbf{R}$. Note that we have assumed that $\mathbf{s}$ and $\mathbf{R}$ have equivalent sampling rates – in practice, either the neural response or the stimulus must be resampled to the lower of the two native sampling rates. To implement the convolution, it will be convenient to define a Toeplitz matrix ${}^{*}\mathbf{s}$ with column and row indices τ and t: $({}^{*}\mathbf{s})_{\tau,t} = s(t-\tau)$, where $s(t)$, are the elements of vector $\mathbf{s}$ and, $\tau = 1\ldots Q$, denotes the taps of the applied temporal filter. Elements prior to the first sample can be set to zero assuming that there is no stimulation prior to the start. This matrix has dimensions $Q{\times}L$. With this we can now define the temporally filtered stimulus and spatially filtered response:

$$\begin{aligned}\mathbf{u} &= \mathbf{h}^T {}^*\mathbf{s},\\ \mathbf{v} &= \mathbf{w}^T\mathbf{R}.\end{aligned} \tag{11}$$

The temporal and spatial filters that maximize the correlation between $\mathbf{u}$ and $\mathbf{v}$ are given by CCA, which provides a set of components encompassing multiple filter pairs $\{\mathbf{h}_k, \mathbf{w}_k, k = 1\ldots K\}$ as the following eigenvectors (Borga, 1998; Hotelling, 1936):

$$\begin{aligned}({}^*\mathbf{s}{}^*\mathbf{s}^\mathbf{T})^{-1}{}^*\mathbf{s}\mathbf{R}^T(\mathbf{R}\mathbf{R}^T)^{-1}\mathbf{R}{}^*\mathbf{s}^\mathbf{T}\mathbf{h}_k &= \rho_k\mathbf{h}_k,\\ (\mathbf{R}\mathbf{R}^T)^{-1}\mathbf{R}{}^*\mathbf{s}^\mathbf{T}({}^*\mathbf{s}{}^*\mathbf{s}^\mathbf{T})^{-1}{}^*\mathbf{s}\mathbf{R}^T\mathbf{w}_k &= \rho_k\mathbf{w}_k.\end{aligned} \tag{12}$$

The number of components K is limited by the rank of the data: $K = min(rank({}^*\mathbf{s}), \mathbf{rank}(\mathbf{R}))$. The maximal correlation of ρ_1 is achieved by projecting the stimulus onto $\mathbf{h}_1$ and the neural response onto $\mathbf{w}_1$. Subsequent components $(\mathbf{h}_k, \mathbf{w}_k)$, $k = 2,\ldots,K$, yield projections temporally uncorrelated with previous ones and progressively lower correlations, such that $\rho_1 > \rho_2 > \ldots\rho_K$.

For practical purposes, it is worth noting that Eq. (12) is the conventional CCA solution which has been implemented in many toolboxes. In particular, one can simply execute the canoncorr function in MATLAB with matrices $\mathbf{R}$ and ${}^*\mathbf{s}$ as inputs. However, in most cases these toolboxes will not have implemented regularization, as discussed next.

Regularization

While the hybrid encoding-decoding model will generally possess fewer parameters than conventional encoding or decoding, it may still be beneficial to regularize the solutions. In particular, CCA inverts the covariance matrices of the neural response and the stimulus ($\mathbf{R}\mathbf{R}^T$ and ${}^*\mathbf{s}{}^*\mathbf{s}^\mathbf{T}$ respectively). Prior to inversion of these matrices, it is important to limit dimensions with small eigenvalues that are dominated by noise. To this end one can substitute the inverse of the covariance matrix $\mathbf{C}$ with:

$$\mathbf{C}^{-1} \leftarrow \mathbf{B}[\Lambda^{-1}]_J\mathbf{B}^T, \tag{13}$$

where $\mathbf{B}$ is a matrix of eigenvectors of $\mathbf{C}$ sorted in descending order of associated eigenvalues, and $[\Lambda^{-1}]_J$ a diagonal matrix with the corresponding eigenvalues inverted and set to zero for all dimensions beyond J. Decreasing J increases the strength of regularization. Here we selected the value of J (i.e., 10) as the knee point of the eigenvalue spectrum for both neural response and stimulus. Importantly, none of the results reported in the main text depended critically on this choice.

Spatial and temporal response

To visualize the spatial distribution of neural activity associated with each component, it is conventional to use the "forward model" formalism (Parra et al., 2005; Haufe et al., 2014). The forward model is defined as the linear mapping that best recovers the neural response $\mathbf{R}$ from the decoded response $\mathbf{V}$ in a least-squares sense, namely

$$\mathbf{A}_r = (\mathbf{V}\mathbf{V}^T)^{-1}\mathbf{V}\mathbf{R}^T. \tag{14}$$

where $\mathbf{V} = \mathbf{W}^T\mathbf{R}$ is the matrix of decoded neural responses, $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2\ldots\mathbf{w}_K]$ is a matrix of K CCA-derived spatial filters, and the corresponding forward models are the columns of matrix $\mathbf{A}_r = [\mathbf{a}_1, \mathbf{a}_2\ldots\mathbf{a}_K]$. The kth column $\mathbf{a}_k$ reflects the spatial mapping from the neural activity $\mathbf{v}_k$ (extracted to correlate with the stimulus) to the scalp sensors.

This forward model is equal (up to a scaling of each component) to the "spatial response" defined here as the linear mapping that best recovers (in a least-squares sense) the neural response $\mathbf{R}$ from the temporally filtered stimulus $\mathbf{U}$:

$$\mathbf{A}_s = (\mathbf{U}\mathbf{U}^T)^{-1}\mathbf{U}\mathbf{R}^T. \tag{15}$$

where $\mathbf{H}$ and $\mathbf{V}$ denote matrices composed of the vectors $\mathbf{h}_k$, and $\mathbf{v}_k$, respectively. The proportionality of the forward model and this spatial response follows from the fact that both $\mathbf{U}\mathbf{U}^T$ and $\mathbf{V}\mathbf{V}^T$ are diagonal matrices, and that $\mathbf{U}\mathbf{R}^T \propto \mathbf{V}\mathbf{R}^T$, where $\propto$ indicates that both sides are equal up to a diagonal scaling matrix (see Eqs. 4.31 and 4.26 in Borga (1998) respectively). Therefore, $\mathbf{A}_r \propto \mathbf{V}\mathbf{R}^T \propto \mathbf{U}\mathbf{R}^T \propto \mathbf{A}_s$.

In total, $\mathbf{H}$ is the "temporal response" and $\mathbf{A}$ is the "spatial response" which together best recover the neural response from the stimulus with the following rank-K linear estimate:

$$\widehat{\mathbf{R}} = \mathbf{A}^T\mathbf{H}^T{}^*\mathbf{s}. \tag{16}$$

Conventional encoding and decoding

For comparison and reference we provide here the equations for the encoding and decoding approaches. To implement the temporal correlation (filtering) in Eq. (6) of the main text we define a Hankel matrix with column and row indices τ and t: $({}^\star\mathbf{r})_{\tau,t} = r(t+\tau)$, and a block-Hankel matrix that concatenates the responses for all sensors, ${}^\star\mathbf{R} = [{}^\star\mathbf{r}_\mathbf{1}^\mathbf{T},\ldots,{}^\star\mathbf{r}_\mathbf{D}^\mathbf{T}]^T$. With that we can write the encoded response and decoded stimulus as:

$$\hat{r}_i = \mathbf{h}_i^T{}^*\mathbf{s}, \tag{17}$$

$$\hat{s} = \mathbf{w}^T{}^\star\mathbf{R}. \tag{18}$$

The best encoding and decoding filters (Eqs. (5) and (6) in the main text) are unique and are given by the conventional least-squares estimates:

$$\mathbf{h}_i = ({}^*\mathbf{s}{}^*\mathbf{s}^\mathbf{T})^{-1}{}^*\mathbf{s}\mathbf{r}_i^T, \tag{19}$$

$$\mathbf{w} = ({}^\star\mathbf{R}{}^\star\mathbf{R}^\mathbf{T})^{-1}{}^\star\mathbf{R}\mathbf{s}^T. \tag{20}$$

Note the similarities of these equations to the CCA equations (12).

References

Akaho, Shotaro, 2001. A kernel method for canonical correlation analysis. In: Proceedings of the international meeting of the psychometric society (IMPS). Berlin: Springer.

Andrew, Galen, Arora, Raman, Bilmes, Jeff A, Livescu, Karen, 2013. Deep canonical correlation analysis. In: Proceedings of the ICML. (3), pp. 1247–1255.

Bialek, William, Rieke, Fred, Van Steveninck, RR de Ruyter, Warland, David, 1991. Reading a neural code. Science 252 (5014), 1854–1857.

Borga, Magnus, 1998. Learning multidimensional signal processing (Ph.D. thesis). Linköping University.

Butts, Daniel A., Weng, Chong, Jin, Jianzhong, Yeh, Chun-I., Lesica, Nicholas A., Alonso, Jose-Manuel, Garrett, B.Stanley, 2007. Temporal precision in the neural code and the timescales of natural vision. Nature 449 (7158), 92–95.

Bießmann, Felix, Meinecke, Frank C., Gretton, Arthur, Rauch, Alexander, Rainer, Gregor, Logothetis, Nikos K., Müller, Klaus-Robert, 2010. Temporal kernel cca and its application in multimodal neuronal data analysis. Mach. Learn. 79 (1–2), 5–27.

Callan, Daniel E., Callan, Akiko M., Kroos, Christian, Vatikiotis-Bateson, Eric, 2001. Multimodal contribution to speech perception revealed by independent component analysis: a single-sweep eeg case study. Cogn. Brain Res. 10 (3), 349–353.

Cohen, Samantha S, Parra, Lucas C, 2016. Memorable audiovisual narratives synchronize sensory and supramodal neural responses. eneuro, 3 (6): ENEURO–0203.

Dayan, Peter, Abbott, Laurence F., 2001. Theoretical Neuroscience 806. MIT Press, Cambridge, MA.

de Cheveigné, Alain, Parra, Lucas C., 2014. Joint decorrelation, a versatile tool for multichannel data analysis. Neuroimage 98, 487–505.

Ding, Nai, Simon, Jonathan Z., 2012a. Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. J. Neurophysiol. 107 (1), 78–89.

Ding, Nai, Simon, Jonathan Z, 2012b. Emergence of neural encoding of auditory objects while listening to competing speakers. In: Proceedings of the National Academy of Sciences. 109 (29), pp. 11854–11859.

Dmochowski, Jacek P., Sajda, Paul, Dias, Joao, Lucas, C.Parra, 2012. Correlated components of ongoing eeg point to emotionally laden attention-a possible marker of engagement? Front. Human. Neurosci. 6 (112).

Dmochowski, Jacek P., Bezdek, Matthew A., Abelson, Brian P., Johnson, John S., Schumacher, Eric H., Lucas, C.Parra, 2014. Audience preferences are predicted by temporal reliability of neural processing. Nat. Commun. 5.

Eyherabide, Hugo Gabriel, Samengo, Inés, 2013. When and why noise correlations are important in neural decoding. J. Neurosci. 33 (45), 17921–17936.

Friman, Ola, Cedefamn, Jonny, Lundberg, Peter, Borga, Magnus, Knutsson, Hans, 2001. Detection of neural activity in functional mri using canonical correlation analysis. Magn. Reson. Med. 45 (2), 323–330.

Friston, Karl J., Holmes, Andrew P., Worsley, Keith J., Poline, J.-P., Frith, Chris D., Frackowiak, Richard SJ, 1994. Statistical parametric maps in functional imaging: a general linear approach. Hum. Brain Mapp. 2 (4), 189–210.

Friston, Karl, Chu, Carlton, Mourão-Miranda, Janaina, Hulme, Oliver, Rees, Geraint, Penny, Will, Ashburner, John, 2008. Bayesian decoding of brain images. NeuroImage 39 (1), 181–205.

Fujiwara, Yusuke, Miyawaki, Yoichi, Kamitani, Yukiyasu, 2013. Modular encoding and decoding models derived from bayesian canonical correlation analysis. Neural Comput. 25 (4), 979–1005.

Gerson, Adam D., Parra, Lucas C., Sajda, Paul, 2006. Cortically coupled computer vision for rapid image search. IEEE Trans. Neural Syst. Rehabil. Eng. 14 (2), 174–179.

Golumbic, Elana M.Zion, Ding, Nai, Bickel, Stephan, Lakatos, Peter, Schevon, Catherine A., McKhann, Guy M., Goodman, Robert R., Emerson, Ronald, Mehta, Ashesh D., Simon, Jonathan Z., et al., 2013. Mechanisms underlying selective neuronal tracking of attended speech at a cocktail party. Neuron 77 (5), 980–991.

Groen, Iris I.A., Ghebreab, Sennay, Prins, Hielke, Lamme, Victor A.F., Scholte, H. Steven, 2013. From image statistics to scene gist: evoked neural activity reveals transition from low-level natural image structure to scene category. J. Neurosci. 33 (48), 18814–18824.

Hasson, Uri, Nir, Yuval, Levy, Ifat, Fuhrmann, Galit, Malach, Rafael, 2004. Intersubject synchronization of cortical activity during natural vision. Science 303 (5664), 1634–1640.

Hasson, Uri, Furman, Orit, Clark, Dav, Dudai, Yadin, Davachi, Lila, 2008a. Enhanced intersubject correlations during movie viewing correlate with successful episodic encoding. Neuron 57 (3), 452–462.

Hasson, Uri, Yang, Eunice, Vallines, Ignacio, Heeger, David J., Rubin, Nava, 2008b. A hierarchy of temporal receptive windows in human cortex. J. Neurosci. 28 (10), 2539–2550.

Hasson, Uri, Landesman, Ohad, Knappmeyer, Barbara, Vallines, Ignacio, Rubin, Nava, Heeger, David J., 2008c. Neurocinematics: the neuroscience of film. Projections 2 (1), 1–26.

Haufe, Stefan, Meinecke, Frank, Görgen, Kai, Dähne, Sven, Haynes, John-Dylan, Blankertz, Benjamin, Bießmann, Felix, 2014. On the interpretation of weight vectors of linear models in multivariate neuroimaging. Neuroimage 87, 96–110.

Haxby, James V., Connolly, Andrew C., Guntupalli, J.Swaroop, 2014. Decoding neural representational spaces using multivariate pattern analysis. Annu. Rev. Neurosci. 37, 435–456.

Honey, Christopher J., Thesen, Thomas, Donner, Tobias H., Silbert, Lauren J., Carlson, Chad E., Devinsky, Orrin, Doyle, Werner K., Rubin, Nava, Heeger, David J., Hasson, Uri, 2012. Slow cortical dynamics and the accumulation of information over long timescales. Neuron 76 (2), 423–434.

Horikawa, Tomoyasu, Tamaki, Masako, Miyawaki, Yoichi, Kamitani, Yukiyasu, 2013. Neural decoding of visual imagery during sleep. Science 340 (6132), 639–642.

Horn, Berthold K, Schunck, Brian G, 1981. Determining optical flow. In: Proceedings of the 1981 Technical symposium east. International Society for Optics and Photonics. pp. 319–331.

Hotelling, Harold, 1936. Relations between two sets of variates. Biometrika 28 (3/4), 321–377.

Huang, Yu, Parra, Lucas C., Haufe, Stefan, 2016. The new york heada precise standardized volume conductor model for eeg source localization and tes targeting. NeuroImage 140, 150–162.

Huth, Alexander G., Shinji Nishimoto, Vu, An T., Gallant, Jack L., 2012. A continuous semantic space describes the representation of thousands of object and action categories across the human brain. Neuron 76 (6), 1210–1224.

Huth, Alexander G., de Heer, Wendy A., Griffiths, Thomas L., Theunissen, Frédéric E., Gallant, Jack L., 2016. Natural speech reveals the semantic maps that tile human cerebral cortex. Nature 532 (7600), 453–458.

Kamitani, Yukiyasu, Tong, Frank, 2005. Decoding the visual and subjective contents of the human brain. Nat. Neurosci. 8 (5), 679–685.

Kay, Kendrick N., Naselaris, Thomas, Prenger, Ryan J., Gallant, Jack L., 2008. Identifying natural images from human brain activity. Nature 452 (7185), 352–355.

Ki, Jason J., Kelly, Simon P., Parra, Lucas, 2016. Attention strongly modulates reliability of neural responses to naturalistic narrative stimuli. J. Neurosci. 36 (10), 3092–3101.

Koskinen, Miika, Viinikanoja, Jaakko, Kurimo, Mikko, Klami, Arto, Kaski, Samuel, Hari, Riitta, 2013. Identifying fragments of natural speech from the listener's meg signals. Hum. Brain Mapp. 34 (6), 1477–1489.

Kriegeskorte, Nikolaus, Goebel, Rainer, Bandettini, Peter, 2006. Information-based functional brain mapping. In: Proceedings of the National academy of Sciences of the United States of America. 103 (10), pp. 3863–3868.

Lahnakoski, Juha M., Glerean, Enrico, Jääskeläinen, Iiro P., Hyönä, Jukka, Hari, Riitta, Sams, Mikko, Nummenmaa, Lauri, 2014. Synchronous brain activity across individuals underlies shared psychological perspectives. NeuroImage 100, 316–324.

Lakatos, Peter, O'Connell, Monica N., Barczak, Annamaria, Mills, Aimee, Javitt, Daniel C., Schroeder, Charles E., 2009. The leading sense: supramodal control of neurophysiological context by attention. Neuron 64 (3), 419–430.

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

Lalor, Edmund C., Pearlmutter, Barak A., Reilly, Richard B., McDarby, Gary, Foxe, John J., 2006. The vespa: a method for the rapid estimation of a visual evoked potential. Neuroimage 32 (4), 1549–1561.

Lankinen, K., Saari, J., Hari, R., Koskinen, M., 2014. Intersubject consistency of cortical meg signals during movie viewing. NeuroImage 92, 217–224.

Lerner, Yulia, Honey, Christopher J., Silbert, Lauren J., Hasson, Uri, 2011. Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. J. Neurosci. 31 (8), 2906–2915.

Di Liberto, Giovanni M, James, A O'Sullivan, Lalor, Edmund C, 2015. Low-frequency cortical entrainment to speech reflects phoneme-level processing. Curr. Biol. 25 (19), pp. 2457–2465.

Luo, Huan, Poeppel, David, 2007. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron 54 (6), 1001–1010.

Mesgarani, Nima, Chang, Edward F., 2012. Selective cortical representation of attended speaker in multi-talker speech perception. Nature 485 (7397), 233–236.

Mesgarani, Nima, Cheung, Connie, Johnson, Keith, Chang, Edward F., 2014. Phonetic feature encoding in human superior temporal gyrus. Science 343 (6174), 1006–1010.

Miyawaki, Yoichi, Uchida, Hajime, Yamashita, Okito, Sato, Masa-aki, Morito, Yusuke, Tanabe, Hiroki C., Sadato, Norihiro, Kamitani, Yukiyasu, 2008. Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. Neuron 60 (5), 915–929.

Monti, Martin M., 2011. Statistical analysis of fmri time-series: a critical review of the glm approach. Front. Human. Neurosci. 5, article 28.

Nandy, Rajesh R., Cordes, Dietmar, 2003. Novel nonparametric approach to canonical correlation analysis with applications to low cnr functional mri data. Magn. Reson. Med. 50 (2), 354–365.

Naselaris, Thomas, Prenger, Ryan J., Kay, Kendrick N., Oliver, Michael, Gallant, Jack L., 2009. Bayesian reconstruction of natural images from human brain activity. Neuron 63 (6), 902–915.

Naselaris, Thomas, Kay, Kendrick N., Nishimoto, Shinji, Gallant, Jack L., 2011. Encoding and decoding in fmri. Neuroimage 56 (2), 400–410.

Nishimoto, Shinji, Vu, An T., Naselaris, Thomas, Benjamini, Yuval, Yu, Bin, Gallant, Jack L., 2011. Reconstructing visual experiences from brain activity evoked by natural movies. Curr. Biol. 21 (19), 1641–1646.

Norman, Kenneth A., Polyn, Sean M., Detre, Greg J., Haxby, James V., 2006. Beyond mind-reading: multi-voxel pattern analysis of fmri data. Trends Cogn. Sci. 10 (9), 424–430.

O'Connell, Redmond G., Dockree, Paul M., Robertson, Ian H., Bellgrove, Mark A., Foxe, John J., Kelly, Simon P., 2009. Uncovering the neural signature of lapsing attention: electrophysiological signals predict errors up to 20s before they occur. J. Neurosci. 29 (26), 8604–8611.

O'Sullivan, James A, Power, Alan J, Mesgarani, Nima, Rajaram, Siddharth, Foxe, John J, Shinn-Cunningham, Barbara G, Slaney, Malcolm, Shamma, Shihab A, Lalor, Edmund C, 2014. Attentional selection in a cocktail party environment can be decoded from single-trial eeg. Cereb. Cortex, pp. bht355.

O'Sullivan, James A., Power, Alan J., Mesgarani, Nima, Rajaram, Siddharth, Foxe, John J., Shinn-Cunningham, Barbara G., Slaney, Malcolm, Shamma, Shihab A., Lalor, Edmund C., 2015. Attentional selection in a cocktail party environment can be

decoded from single-trial eeg. Cereb. Cortex 25 (7), 1697–1706.
Parra, Lucas C., Spence, Clay D., Gerson, Adam D., Sajda, Paul, 2005. Recipes for the linear analysis of eeg. Neuroimage 28 (2), 326–341.
Pasley, Brian N., David, Stephen V., Mesgarani, Nima, Flinker, Adeen, Shamma, Shihab A., Crone, Nathan E., Knight, Robert T., Chang, Edward F., et al., 2012. Reconstructing speech from human auditory cortex. PLoS-Biol. 10 (1), 175.
Pereira, Francisco, Mitchell, Tom, Botvinick, Matthew, 2009. Machine learning classifiers and fmri: a tutorial overview. Neuroimage 45 (1), S199–S209.
Picton, T.W., Hillyard, S.A., 1974. Human auditory evoked potentials. ii: effects of attention. Electroencephalogr. Clin. Neurophysiol. 36, 191–200.
Poulsen, Andreas Trier, Kamronn, Simon, Dmochowski, Jacek, Parra, Lucas C., Hansen, Lars Kai, 2017. EEG in the classroom: synchronised neural recordings during video presentation. Sci. Rep. 7.
Power, Alan J., Foxe, John J., Forde, Emma-Jane, Reilly, Richard B., Lalor, Edmund C., 2012. At what time is the cocktail party? A late locus of selective attention to natural speech. Eur. J. Neurosci. 35 (9), 1497–1503.
Raij, Tommi, Uutela, Kimmo, Hari, Riitta, 2000. Audiovisual integration of letters in the human brain. Neuron 28 (2), 617–625.
Salisbury, Dean F., Rutherford, Bret, Shenton, Martha E., McCarley, Robert W., 2001. Button-pressing affects p300 amplitude and scalp topography. Clin. Neurophysiol. 112 (9), 1676–1684.
Saupe, Katja, Schröger, Erich, Andersen, Søen K., Müller, Matthias M., 2009. Neural mechanisms of intermodal sustained selective attention with concurrently presented auditory and visual stimuli. Front. Human. Neurosci. 3, 58.
Silbert, Lauren J, Honey, Christopher J, Simony, Erez, Poeppel, David, Hasson, Uri, 2014. Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. In: Proceedings of the National Academy of Sciences, 111 (43), pp. E4687–E4696.
Stanley, Garrett B., Li, Fei F., Dan, Yang, 1999. Reconstruction of natural scenes from ensemble responses in the lateral geniculate nucleus. J. Neurosci. 19 (18), 8036–8042.
Stephens, Greg J, Silbert, Lauren J, Hasson, Uri, 2010. Speaker-listener neural coupling underlies successful communication. In: Proceedings of the National Academy of Sciences, 107 (32), pp. 14425–14430.
Theiler, James, Eubank, Stephen, Longtin, André, Galdrikian, Bryan, Farmer, J. Doyne, 1992. Testing for nonlinearity in time series: the method of surrogate data. Phys. D: Nonlinear Phenom. 58 (1–4), 77–94.
Thirion, Bertrand, Duchesnay, Edouard, Hubbard, Edward, Dubois, Jessica, Poline, Jean-Baptiste, Lebihan, Denis, Dehaene, Stanislas, 2006. Inverse retinotopy: inferring the visual content of images from brain activation patterns. Neuroimage 33 (4), 1104–1116.
Walz, Jennifer M., Goldman, Robin I., Carapezza, Michael, Muraskin, Jordan, Brown, Truman R., Sajda, Paul, 2013. Simultaneous eeg-fmri reveals temporal evolution of coupling between supramodal cortical attention networks and the brainstem. J. Neurosci. 33 (49), 19212–19222.
Wang, Weiran, Arora, Raman, Livescu, Karen, Bilmes, Jeff A, 2015. Unsupervised learning of acoustic features via deep canonical correlation analysis. In: Proceedings of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, pp. 4590–4594.
Warland, David K., Reinagel, Pamela, Meister, Markus, 1997. Decoding visual information from a population of retinal ganglion cells. J. Neurophysiol. 78 (5), 2336–2350.
Worsley, Keith J., Poline, Jean-Baptiste, Friston, Karl J., Evans, A.C., 1997. Characterizing the response of pet and fmri data using multivariate linear models. NeuroImage 6 (4), 305–319.
Yamashita, Okito, Sato, Masa-aki, Yoshioka, Taku, Tong, Frank, Kamitani, Yukiyasu, 2008. Sparse estimation automatically selects voxels relevant for the decoding of fmri activity patterns. NeuroImage 42 (4), 1414–1429.