Decoding in the fourth dimension: Classification of temporal patterns and their generalization across locations

Alejandro Santos-Mayo[1], Faith Gilbert[1], Laura Ahumada[1], Caitlin Traiser[1], Hannah Engle[1], Christian Panitz[2], Mingzhou Ding[3], & Andreas Keil[1]

[1]*Department of Psychology, University of Florida, Gainesville, FL, United States*

[2]*Department of Psychology, University of Bremen, Bremen, Germany*

[3]*J. Crayton Pruitt Family Department of Biomedical Engineering, University of Florida, Gainesville, FL, United States*

Abstract

Neuroscience research has increasingly used decoding techniques, in which multivariate statistical methods identify patterns in neural data that allow the classification of experimental conditions or participant groups. Typically, the features used for decoding are spatial in nature, including voxel patterns and electrode locations. However, the strength of many neurophysiological recording techniques such as electroencephalography or magnetoencephalography is in their rich temporal, rather than spatial, content. The present report proposes a new decoding method that relies on the time information contained in neural time series. This information is then used in a subsequent step, generalization across location (GAL), which characterizes the relationship between sensor locations based on their ability to cross-decode. Two datasets are used to demonstrate usage of this method, referred to as time-GAL, involving (1) event-related potentials in response to affective pictures and (2) steady-state visual evoked potentials in response to aversively conditioned grating stimuli. In both cases, experimental conditions were successfully decoded based on the temporal features contained in the neural time series. Cross-decoding occurred in regions known to be involved in visual and affective processing. We conclude that the time-GAL approach holds promise for analyzing neural time series from a wide range of paradigms and measurement domains providing an assumption-free method to quantifying differences in temporal patterns of neural information processing and whether these patterns are shared across sensor locations.

Keywords: Decoding, temporal patterns, electrophysiology, multivariate pattern analysis

Introduction

The development of human neuroimaging tools such as functional magnetic resonance imaging (fMRI) and magneto- and electroencephalography (M/EEG) has transformed the study of human brain function. Traditionally, these measurements have been analyzed using mass univariate statistics (Friston et al., 1994), in which a linear statistical model is applied to a given time point, voxel, or electrode. Because of the univariate nature of this approach, it disregards information regarding spatial or temporal patterns (Mahmoudi et al., 2012). Thus, multivariate pattern analysis (MVPA) has emerged as an alternative tool for discriminating the neural activity patterns related to a given stimulus, experimental condition, or participant (Carrasco et al., 2024; Haxby, 2012).

Multivariate decoding techniques capitalize on the information found in neural patterns to characterize the multifaceted properties that define a given brain process (Peelen & Downing, 2023). Also known as decoding methods, these multivariate techniques were originally introduced for fMRI data (Haxby, 2012; Haxby et al., 2001), and their use has steadily increased in recent years, becoming a standard procedure in this field (Peelen & Downing, 2023; Rivolta et al., 2014). Meanwhile, in the field of M/EEG research, decoding approaches are also increasingly used (Grootswagers et al., 2017) with several standard procedures emerging (Ashton et al., 2022; Bae & Luck, 2019; Bo et al., 2022). Generally, these standard procedures use spatial information to decode the underlying neural activity patterns, paralleling approaches used with fMRI data (see Figure 1 left). However, the resolution of spatial information is typically reduced for electrophysiological tools like EEG in comparison with fMRI. It is widely accepted that the primary strength of EEG and MEG recordings lies in their rich temporal information, rather than in their spatial distribution. Specifically, human electrophysiology reflects the time series of neural population activity arising from post-synaptic events, subject to volume conduction and other distortions when measured from the scalp. The resulting waveforms contain valuable information

about the neural response and/or oscillatory rhythms related to a specific experimental condition or stimulus (David et al., 2006).

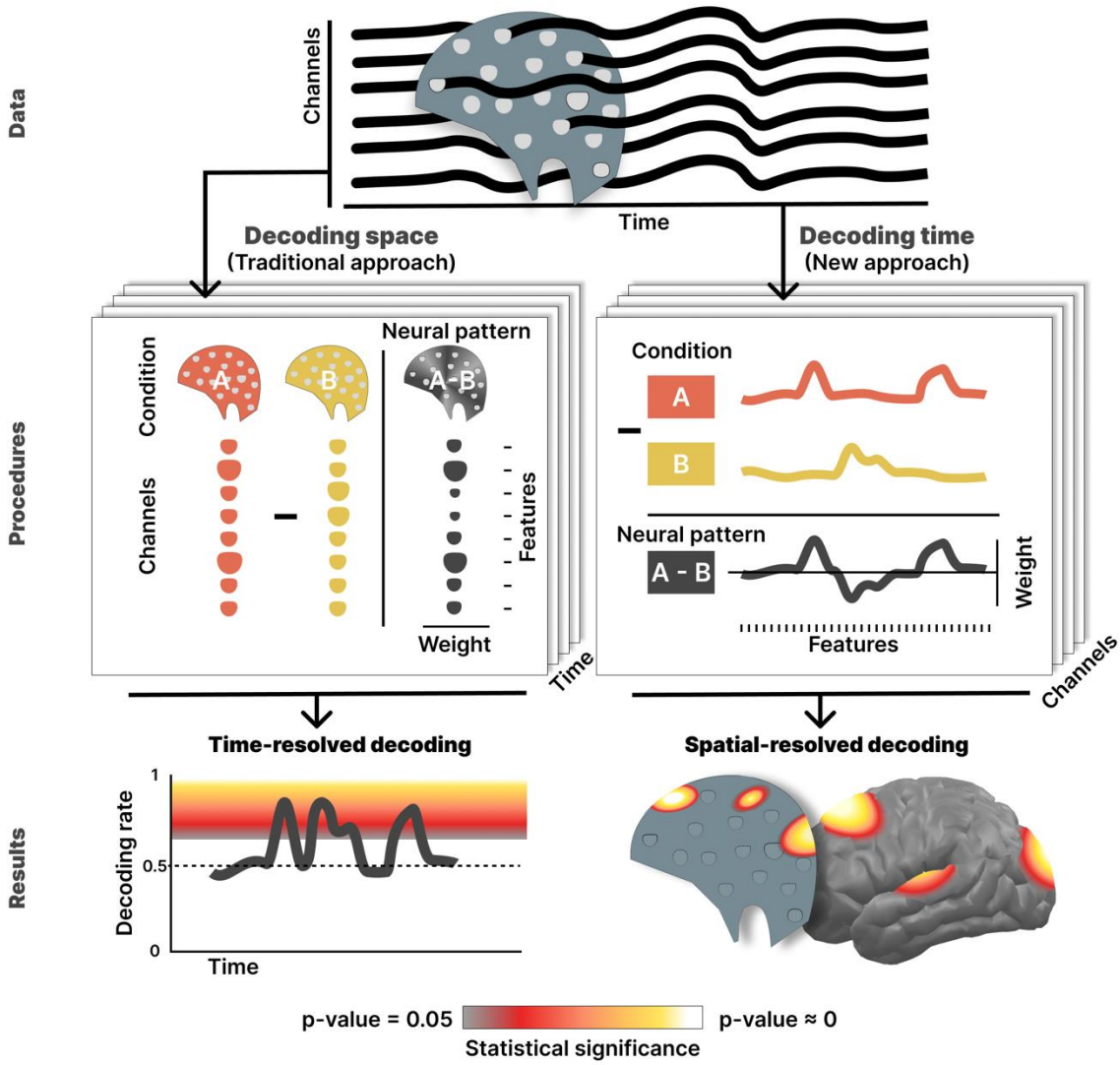

Figure 1: Flowchart describing how MVPA-based decoding techniques can be carried out following the traditional approach (**left**) or the present in this work (**right**). Data from electrophysiological recordings (**top**) offers both spatial and temporal information in the form of channels or time-series, respectively, that can be used for decoding. Thus, each procedure (**middle**) leverages one of the dimensions using its data to fit a classifier, repeating this operation over the other dimension. Specifically, the traditional approach employs channels or sensors as features; the time-GAL approach uses temporal features. Finally, the

results (**bottom**) for both the traditional (space-based) and the time-GAL method show the decoding performance across time or locations, respectively.

In the present study, we systematically explore the suitability of applying MVPA algorithms to the temporal dimension of EEG signals, taking advantage of the high resolution of EEG data in this dimension. Specifically, we use all time points of each trial time series as features that enter a multivariate classifier model (see Figure 1, right), in contrast to the spatial information that has previously been used in the literature. Then, we repeat this procedure for each spatial point (e.g. electrodes in the case of EEG) to obtain a spatially resolved depiction of scalp regions in which the temporal dynamics related to a condition or stimulus contain overlapping decodable information based on temporal features. In a second step, we adapt a method proposed by King and Dehane (2014) called generalization across time (GAT), in which a classifier algorithm is trained using spatial information for one (training) time point and then tested across a range of different time points. This procedure allows researchers to characterize the temporal stability of neural patterns related to specific conditions or stimuli. The present approach, using the temporal waveform as a feature vector for classification allows us to quantify the extent to which different recording sites share decodable information. This is implemented as a generalization across location (GAL), in which a temporal decoder trained at one location is applied to other locations. By mapping the cross-decoding accuracy across locations, the time-GAL method, offers a multivariate approach to defining the unique topographic patterns in which temporal dynamics are related to a specific stimulus or experimental condition. Finally, another important contribution of the current work consists of the combination of not only spatial, but also temporal information to the decoding output. Classifier models are considered backward models as they extract the information about experimental manipulations from the brain's responses to these manipulations. On the other hand, forward models describe the process in which experimental manipulations and brain biophysics lead to the recorded neurophysiological data. The exact weights of a

classifier-based backward model can only be extracted using a corresponding linear forward model (Haufe et al., 2014). Therefore, we implemented a simple linear forward model based on correlations with the labels to visualize the dynamics of time-GAL decoding and cross-decoding patterns.

In the present study, we demonstrate the usage of the time-GAL approach for analyzing neural time series data, using two EEG experiments. The experiments involved (1) affective picture viewing and (2) aversive generalization conditioning.

Methods

Datasets

To provide a demonstration of the time-GAL procedure, two datasets were used, involving two types of electrophysiological response: event related potentials (ERP) and steady-state visual evoked potentials (ssVEP). ERPs are typically generated by the repeated presentation of transient stimuli while the EEG is recorded, often at temporal intervals of several seconds. Stimulus-locked EEG segments obtained for each presentation are then averaged, resulting in a prototypical voltage time course (David et al., 2006; Di Russo et al., 2002). By contrast, ssVEPs result from rhythmic presentation of visual stimuli (e.g. regularly flickering light), driving an oscillatory neural response at the same temporal frequency as the driving stimulus (Norcia et al., 2015; Wieser et al., 2016).

Data set 1 comes from a picture viewing task in which ERPs were measured from 39 participants (27 women; average age 19.6 years) while passively watching pleasant, neutral, and unpleasant affective pictures from the international affective picture system (IAPS; Lang et al., 2008). The present demonstration focuses on two conditions, pleasant and unpleasant picture viewing. These conditions were selected because they each prompt robust ERP signals but tend to show only small condition differences when compared using univariate methods (Gibney et al.,

2020). 20 pleasant and 20 unpleasant pictures were shown twice each, for a duration of 1 second at the central portion of the screen, resulting in a total of 80 trials per participant, with 40 trials in the pleasant and 40 in the unpleasant condition.

Data set 2 comes from a fear conditioning paradigm in which neutral cues were flickered at 15 Hz, prompting an oscillatory SSVEP response, recorded in 31 participants (22 women; average age 19.23 years). Visual cues consisted of 4 high-contrast circular sinusoidal gratings (orientations: 15°, 35°, 55° and 75°) filtered with a Gaussian envelope (Gabor patches). Each stimulus was presented against a dark gray background in a flickering (15 Hz) mode for 3 seconds. The fear generalization protocol (see Lonsdorf et al., 2017 for a review of aversive conditioning methods) involved two phases: First, in the habituation phase, each of the 4 visual cues was presented 30 times in random order. Later, during the fear acquisition phase, the 4 visual cues were again shown, but one of them (CS+, 15° or 75° randomly assigned per subject) was followed by an aversive loud (88 or 91 SPL dB) white noise (the unconditioned stimulus, US) that started 2 seconds after the onset of the CS for a duration of 1 second, thus co-terminating with the CS. After acquisition, the CS+ orientation typically acquires aversive properties, while the remaining 3 orientations served as generalization stimuli (GS) with increasing dissimilarity relative to the CS+ based on their orientation. Thus, GS1 was the closest stimulus to CS+ in orientation and GS3 was the most dissimilar one.

Procedures of both experiments were approved by the University of Florida institutional review board and informed consent was signed by all participants in accordance with the Declaration of Helsinki.

EEG data acquisition and preprocessing

EEG signals from both datasets were collected using an Electrical Geodesics (EGI, Eugene, OR) system with a 128-channel HydroCel electrode net, and continuously digitized at 500 Hz, referenced to the vertex sensor (Cz). Electrode impedances were kept below 50 kΩ.

Online (low-pass filter, 3 dB point at 170 Hz; high-pass filter, 3 dB point at 0.05 Hz) and offline (low-pass filter, 40 Hz, 18th order, 3 dB point at 0.05 Hz) Butterworth filters were applied to the data. Then, the signal was segmented in epochs including 0.2 s pre-stimulus and 1 s post-stimulus for the ERP and 1 s pre-stimulus, and 2 s post-stimulus for the ssVEP datasets. Statistical Correction of Artifacts in Dense Array Studies (SCADS) procedure (Junghofer et al., 2000) was employed for artifact rejection of bad channels and epochs. Finally, all trials were individually baseline corrected by subtracting the voltage activity average in the pre-stimulus segment of the same trial and ordered in categories pleasant (1037 trials across all participants) and unpleasant (1133 trials) for the ERP dataset and CS+ (548 trials), GS1 (570 trials), GS2 (571 trials), GS3 (578 trials) for habituation and acquisition phases of the ssVEP dataset.

Current source density (CSD)

EEG recordings are characterized by pronounced spatial smoothing and distortion as signals travel from the brain to the extracranial sensor locations, caused by a number of biophysical processes including volume conduction and spatial lowpass filtering. The reduction of shared information is crucial in terms of assessing the spatial generalization of time-based decoders (Keil et al., 2022). Here we applied the current source density (CSD) transformation, an established algorithm for diminishing spatial smearing in the topographical distribution of EEG signals (Nunez et al., 1997). Specifically, we used the implementation by Junghöfer et al (1997) setting the smoothing (regularization) parameter λ to a value of 2.

Decoding analysis for brain response generalization across spatial and time dimensions

Electrophysiological waveform-based classification. In the current report, we first apply (train) a linear discriminant decoder to discriminate two conditions based on temporal features of the electrophysiological signal. Then, in a second step, we quantify the generalization of the temporal decoders across channels. This approach is similar to the generalization across

time (GAT) method (King & Dehaene, 2014), where a spatial decoder is implemented (trained) to discriminate conditions within one or more time points, followed by its application (testing) across a range of different time points. Although GAT sheds light on the decoding time course, the topographic information and the relation between different sensor locations cannot be seen. This report introduces a new method, the generalization across locations (GAL) approach, to address this gap. In contrast to the GAT method, the time-GAL algorithm takes advantage of the temporal information, i.e. the signal waveform, to train a decoder for each channel. Each channel's decoder is then tested at all other channel locations, a process known as cross-decoding. Because this decoding approach aims to infer neural processes from data, it can be considered a backward model (Haufe et al., 2014). In contrast, a forward model is a model that describes how brain activity results in neural data as collected for example by EEG. Because there is a corresponding forward model to each backward model (Haufe et al., 2014), we can employ a forward modeling approach to obtain the information regarding when each topographic channel reflects condition differences.

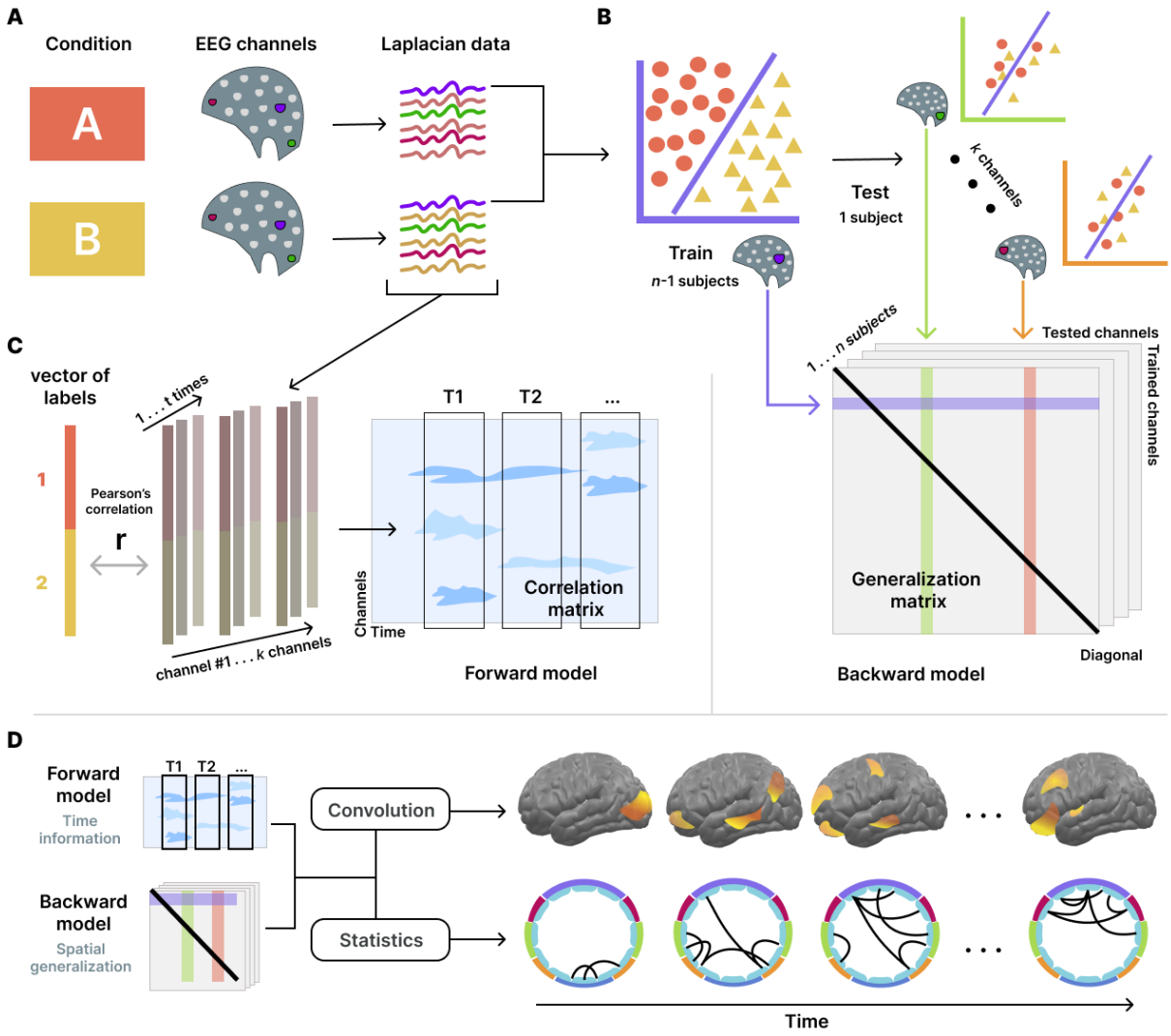

Figure 2: Flowchart describing the decoding methodology based on the temporal and spatial information. **A**: Laplacian EEG data from conditions A and B are used to train one classifier per channel using time points as features. **B**: Next, the generalization matrix (backward model) is calculated by testing every classifier model with data from each of the remaining channels (cross-decoding). **C**: Then, the correlation matrix (forward model) is calculated using Pearson's correlations between the A/B labels and EEG data for each channel and time. **D**: Finally, the combination of both backward and forward models containing the spatial and temporal information, respectively, allows us to visualize the GAL cross-decoding patterns in the brain. We interpret these as similarity or connectivity matrices, akin to correlation-based connectivity matrices used in EEG and fMRI research.

Backward model of the generalization across location. For the GAL analysis, we use data from two conditions, referred to as A and B. In our implementation, each data set consisted of three-dimensional arrays (channels by time points by trials) for each condition and each subject (see Figure 2 top left). For one of *k* channels, we fed the classifier model all time points as features and all single trials from all *n* participants except from one participant (leave-one-out training/testing), as observations. Because the classifier operates at the single trial level, many observations were considered for the training phase of the classifier model, creating pressure on computing resources. Therefore, we compared the performance and efficiency of a linear discriminant analysis (LDA) with a support vector machine (SVM). LDA is proven to be suitable for neural decoding (Mandelkow et al., 2016), although SVM has been more widely used in EEG data (Bae & Luck, 2018, 2019) for averaged ERPs. Similar to Grootswagers et al. (2017), we found comparable decoding rates from both types of models but a significant reduction of time consumption from the LDA algorithm. Because we perform a substantial amount of classifier training and we seek a computationally efficient methodology, we selected LDA as our classifier model to discriminate training trials as coming from conditions A or B. Models were trained using the LDA function built in MATLAB (2023b). Even when the number of trials were unbalanced, all single trials from conditions A and B were employed to avoid randomness in any aleatory selection of trials. Priors were set accordingly. Information contained at single-trials is noisier than evoked information from trial-averages (e.g., ERPs), but single-trial time courses include the non-phase-locked temporal features of the signal, often referred to as “induced”, instead of “evoked” responses (David et al., 2006). More information about the difference between applying the GAL methodology to single-trial or averaged is given in Supplementary 1.

In the second step of the time-GAL method, we applied the classifier model weights trained at a given channel to quantify its decoding performance with data from all *k* channels using the single trials of the remaining subjects as the test data set. Thus, we required the decoder to use feature weights that were applicable across participants. The GAL test phase provides a

classification of trials into A and B that quantifies the accuracy of the model using the decoding accuracy (correct A and B classification) as an index. This procedure was repeated *k* times, training the classifier model with all the remaining channels and testing their weights with all the *k* channels to fill out the decoding accuracy in a generalization matrix of training by testing axes with dimensions *k* by *k* channels (see Figure 2 middle right). Finally, this analysis was computed again for each subject, using the rest of the *n* - 1 participants as training dataset in a leave-one-out approach. Thus, this approach to cross-validation avoids arbitrary decisions (e.g. regarding the number of k-folds) in the training/test data set selection. Finally, this methodology provides us with *n* generalization matrices of channel-by-channel dimension, where *n* corresponds to the number of subjects. This distribution is employed later in the following statistical analyses. Conceptually, if a time-based decoder trained at location k_i also classifies the data at location k_j, then we interpret this as evidence that the temporal dynamics at k_i and k_j contain share decodable information about the conditions—a non-parametric way of describing spatial dependency or connectivity.

Forward model of the temporal contribution. The GAL method allows us to inspect the spatial distribution of the decoding performance between A and B conditions, based on the electrophysiological waveform, or time information. However, this approach offers only a static picture of the connectivity matrix. In the GAL framework, a “connection” indicates that two channels share decodable information. Therefore, it is helpful to visualize what time points in the waveform are most strongly related to the discrimination of conditions. As stated above, it is possible to estimate a forward model to obtain the weights of the LDA features (Haufe et al., 2014). We can address this estimation by measuring the relations between the condition labels and the Laplacian EEG data. To this end, the Pearson’s correlation coefficient between the true vector of A and B labels and all subjects’ trials of electrophysiological activity was calculated for each time point and channel resulting in a channel by time matrix of correlations (see Figure 2 middle left).

Combining decoding topography with the time domain. While the backward model provides spatial information, we use the forward model to visualize the temporal progression of condition differences at the single trial level. Specifically, we combine the information from both models to obtain a temporally weighted representation of the decoding patterns. First, we were interested in a brain topographic visualization of the main contributor areas to the decoding analysis. To obtain this visualization, we used the diagonal vector from the average of the *n* generalization matrices, i.e., the decoding accuracy of each channel's LDA, across participants. This information served to project the decoding rates of the *k* channels into the source space. Then, the decoding-weighted forward model is obtained by timepoint-wise multiplication of the overall decoding vector of *k* spatial values with the absolute values of the *k* by *time* correlation matrix. The absolute value of the correlation matrix is used because it indicates the contribution of a given channel and time point to the discrimination between conditions A and B, regardless of the sign. This combination of the decoding vector (spatial information) and correlation matrix (temporal information) results in another *k* by *time* matrix, which reflects the time-weighted spatial contribution to the decoding. Thus, we obtain one topography that represents the spread of decodable information across the brain surface for every *time* step. EMEGS 2.8 (Peyk et al., 2011) was used to project this information onto a standard cortical surface, for visualization.

Statistical extraction of GAL connectivity patterns over time. For the statistical analysis of GAL data, a mass univariate approach was chosen. Once the *n* generalization matrices were computed, Student's T test against the decoding chance rate were computed for each element in both directions, i.e., above and under the chance level. Since we only have conditions A and B, the decoding chance rate of the classifier is 0.5. The diagonal of the GAL matrix shows the decoding accuracy for the test dataset of the same channel used for classifier training, and thus its values are not expected to be smaller than chance, i.e. 0.5. However, the values for the rest of the generalization matrix, in which test and training datasets came from different channels may well show values below chance level. This would indicate that the data

from one location are systematically located on the other side of the hyperplane spanned by the LDA trained on another location. This may be interpreted as those locations showing orthogonal neural dynamics, more like the other label, resulting in significantly below-chance decoding. Therefore, we used the contrast T test in both directions showing not only those positive connections between regions, but also negative connections (below chance cross-decoding). Each of these analyses resulted in one matrix of channel-by-channel dimensions showing a large number of comparisons. Since connections between different channel locations cannot be considered spatially near, i.e. neighbors, their independence can be assumed, and Bonferroni correction was applied to address the multiple-comparison problem. Also, to obtain only the most meaningful generalization connections, a corrected alpha (α) was utilized (see results). Therefore, only those connections presenting a p-value less than $\frac{\alpha}{k^2}$ where α represents the desired alpha level, were considered statistically significant and used for the description of the connectivity (GAL) pattern.

For visualization, in a final step, we combined the statistically significant GAL results with the statistically significant temporal correlations to illustrate the time-GAL dynamics. Specifically, we only considered the Pearson's correlation significance of those channels that showed statistically significant generalization connections (see above). We used a threshold of $\alpha = 0.01$ to identify temporal correlations (contributions) of each location that were relevant for visualization. This alpha value was not Bonferroni corrected because it was only used as a threshold for visualization. This step allowed us to visualize the relationship between areas throughout the temporal progression of the neural information processing. Thus, we can inspect how different networks or areas are implicated in different time steps, providing a descriptive video of the cognitive processing instead of a static picture of all the GAL pattern.

The above-described methodology is implemented in an open-source MATLAB toolbox. This toolbox can be found in https://github.com/csea-lab/time-GAL and example data obtained in

https://osf.io/q56ns. Also, scripts to calculate the forward model through Pearson's correlation and combine its temporal information with GAL outcomes are included.

Results

The time-GAL decoding methodology uses a classification approach to quantify decodable temporal information in neural dynamics, including how this information generalizes across locations. In the following, we first illustrate how the time-GAL methodology captures information contained in the ERP response to transient visual stimuli. Later, we examine the application of time-GAL to ssVEP responses.

Time-GAL patterns on ERP: Affective pictures presentation

The first dataset consists of CSD-transformed EEG recordings obtained during the presentation of pleasant and unpleasant pictures from a standardized picture database, in a within-participants design. Given this paradigm, we used time-GAL to perform pairwise decoding of the two conditions (pleasant vs unpleasant). The GAL approach was computed and a Bonferroni-corrected α of 0.05 was used as the statistical threshold for the generalization connectivity patterns. Results are shown in Figure 3.

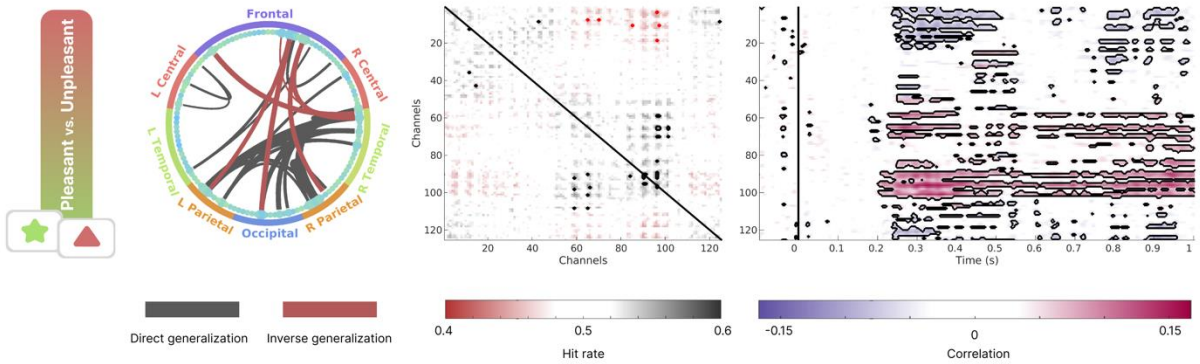

Figure 3. Generalization connectivity patterns and temporal distributions for the affective picture viewing experiment. **Left**: Connectivity graph showing the significant generalization connections between the CSD channels. Black lines refer to connections in which both channels share decodable temporal patterns, while red lines show the opposite pattern (significant below chance decoding). **Middle**: Generalization connectivity matrix showing decoding results (accuracy). The Y axis corresponds to the channel data employed to train the model, and the X axis shows the channel data used for the generalization test. Black and red contours indicate the positive and negative generalization connections, respectively, after a Bonferroni-corrected statistical ($\alpha = 0.05$) comparison against chance (50%). **Right**: Temporal correlation between labels of condition and data showing the positive or negative weights of the decoding model. Black contours indicate the time points whose p-value is below 0.05 for each channel.

Several posterior channels showed significant decoding based on their time dynamics. Similarly, significant generalization connections were more concentrated among posterior areas. It is worth noting that inverse connections (below chance decoding) were observed between anterior and posterior regions, likely representing opposite voltage of dipolar fields when measured at opposite sides of the head. Latencies of correlation between condition labels and channel and time points indicate the emergence of decodable information around 250 ms after trial onset.

Next, we combined the spatial information of the generalization response across the brain with the temporal weights of the model for each channel. Specifically, we multiplied the diagonal of the generalization connectivity matrix with the Pearson's correlation matrix of time weights. Additionally, we determined which generalization connections and temporal weights were statistically significant. As shown in Figure 4, both sources of information shed light on cortical surface regions that contain decodable information (brain surface plots, bottom panels) and how they share decodable information (circular graphs).

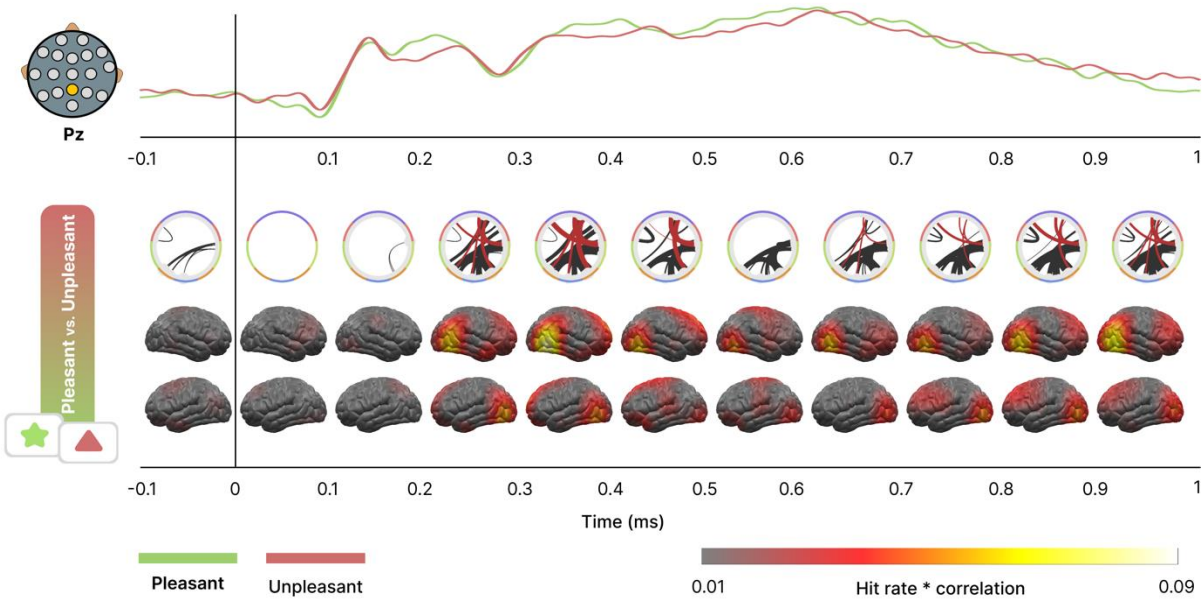

Figure 4. Time course of emotional content decoding and GAL connectivity patterns during affective picture viewing. **Top**: Electrophysiological ERP response to the presentation of different emotional content obtained at the parietal central (Pz) electrode for the three conditions (pleasant, neutral and unpleasant). **Middle**: Circular graphs show the GAL connectivity pattern at each time step for each pairwise condition. The topographies (**Bottom** panel) indicate the multiplication of the decoding rate with the time-varying correlation, showing changes in cross-decoding across the surface of the brain as a function of time.

The time course analysis showed strong inverse decoding between posterior and anterior regions and several direct (positive) generalization connections between right occipital-temporal and parietal zones. The fronto-occipital connectivity pattern was present from 0.2 s to 0.5 s and at the end of the trial from 0.7 s to 1 s, while right occipito-temporal generalization persisted from 0.2 s to the end of the trial. As discussed above, the inverse direction of the GAL results in the antero-posterior connections likely indicates that the voltage patterns found in the occipital and parietal areas are inverted at frontal sensors. By contrast, the right parietal time-GAL patterns are consistent with studies of the late positive potential as mentioned below in the discussion section.

Time-GAL patterns in ssVEP signals: Aversive conditioning of oriented Gabor patches

Next, we applied the timed-GAL approach to ssVEP data from an experiment using flickering visual stimuli. Given the high signal to noise ratio of ssVEPs, we employed a more conservative Bonferroni-corrected α of 0.001, to plot only the most salient generalization connections in the GAL analysis. Results presented below in Figure 5 use data from the acquisition phase of the fear conditioning task, meaning that the CS+ stimulus (a grayscale grating defined by its orientation) became a threat cue due to consistent pairing with an aversive white noise (US).

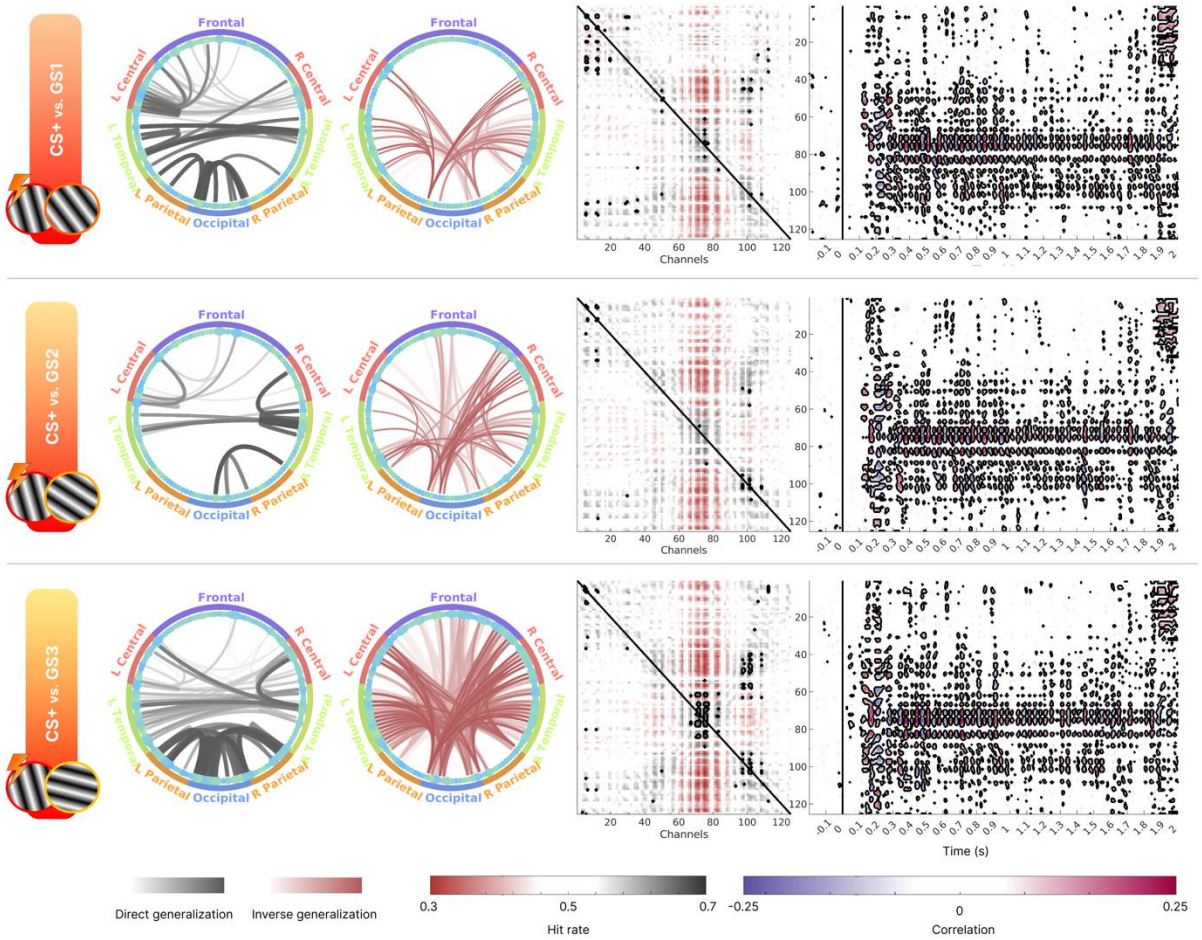

Figure 5. Generalization connectivity patterns during fear conditioning. **Left**: Circular graphs showing the statistically significant generalization connections between 125 CSD channels. Black lines one refers to those connections where both channels show similar decoding temporal patterns, while graph filled with red lines indicate channels showing inversely similar decoding patterns. **Center**: Generalization connectivity matrix showing decoding results (accuracy). Y axis corresponds with the channel data employed to train the model, while X axis refers to the channel data used in the test phase. Black and red contours indicate those direct and inverse generalization connections, respectively, after a Bonferroni-corrected ($\alpha = 0.001$) statistical comparison with chance level (0.5). **Right**: Temporal correlation between labels of condition and data showing the positive or negative weights of the decoding model. Black contours indicate those time points whose p-value is below 0.01 for each channel.

In terms of direct generalization connections (black lines in Figure 5), the first pairwise comparison, CS+ vs. GS1, shows shared decodable information across frontal and central areas along with the expected occipital areas. Interestingly, direct connections were also found between occipital and frontal areas. The next pairwise comparison, CS+ vs. GS2, showed less frontal generalization and general generalization between temporal and central areas. Finally, the CS+ vs. GS3 comparisons, in which GS and CS+ differed the most, showed strong posterior (parietal and occipital areas) generalization connectivity. Regarding the inverse generalization (red lines in Figure 5), a pronounced generalization was observed, driven by occipital areas, as expected when using ssVEPs. In the first case, the comparison between fear cue and the most similar oriented mainly shows an occipito-temporal connection. However, in the case of the GS2, the inverse generalization from the occipital region reaches all the remaining areas. This spread out of generalization appears more potentiated in the case of GS3, remarkably in the frontal areas. Therefore, the results of the inverse GAL point to an increase and spread out of inverse generalization as a function of dissimilarity between the fear cue and the GS. Thereby, differences between conditions in visual areas would be expressed in the opposite direction along the brain and interestingly at the frontal areas.

Taken together, results indicate distinctive involvement and behavior of anterior and posterior areas directly related with the similarity of the visual object with the threat cue. Finally, compared to the above ERP analysis, the ssVEP shows a fluctuating temporal correlation (Figure 5 right) reflecting the oscillation of the ssVEP between positive and negative voltage. This oscillation in weights for the decoding model was expected since wave-form of the trial, i.e. time dimension, was used as observation for training the model and it corresponds with an ssVEP response. In the case of CS+ vs. GS, all the three pairwise showed a similar temporal weight pattern with a preponderance of visual areas and frontal areas before US onset. We combined the weights with the GAL information as shown in Figure 6.

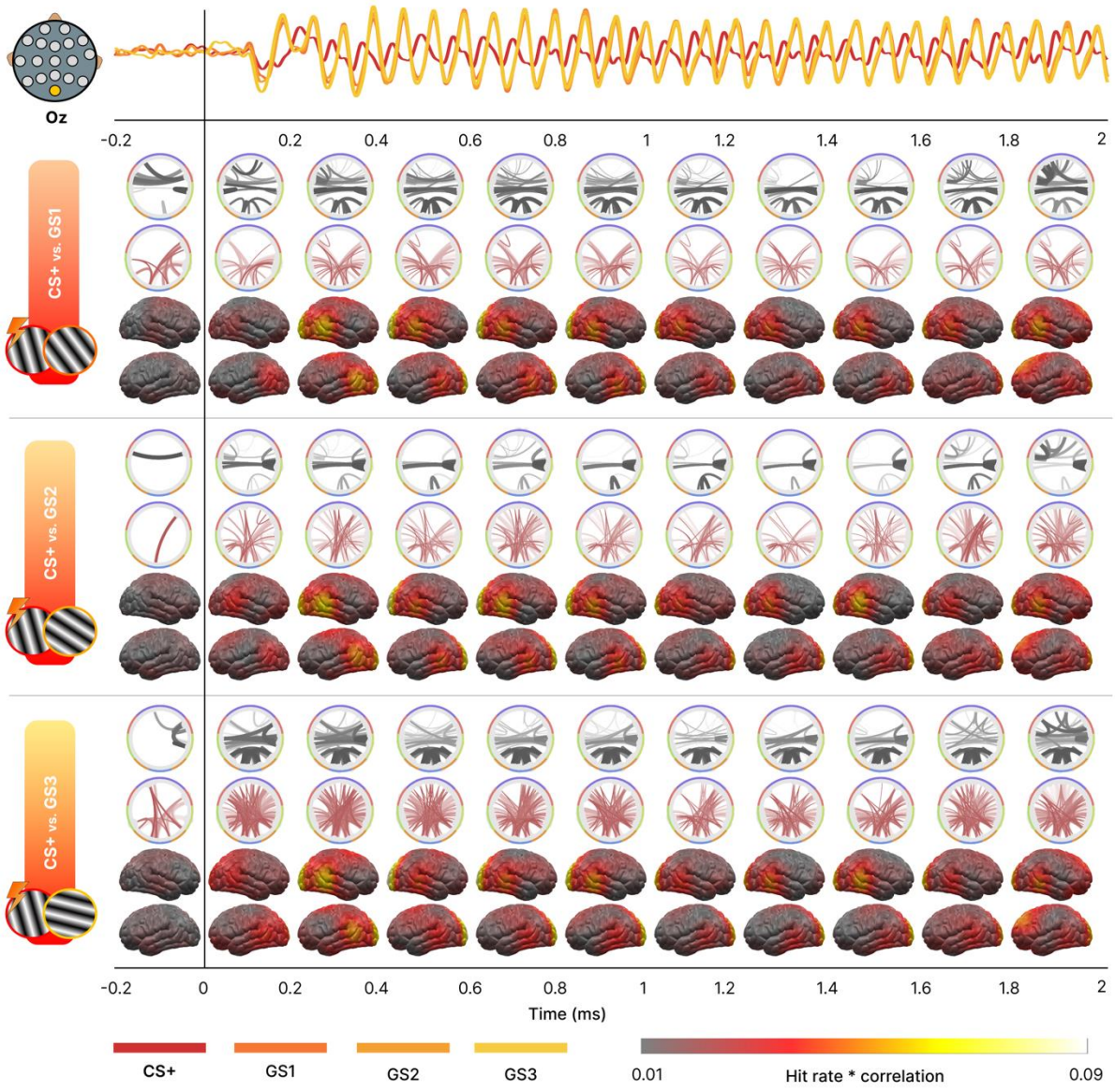

Figure 6. Time course of emotional content decoding and GAL connectivity patterns in fear conditioning. **Top**: Electrophysiological ssVEP response to the presentation of relevant or irrelevant cues obtained at the Occipital central (Oz) electrode for the four conditions (CS+, GS1, GS2, GS3). **Middle**: Circular graphs show the GAL connectivity pattern at each time step for each pairwise condition. For clarity, direct (black lines) and inverse (red lines) generalization across the brain were individually displayed. Brain's projections indicate the convolution of decoding rate by time correlation describing the fluctuation of the fear discrimination process during the trial over the cortex.

Here, we found that the above-mentioned anterior and posterior modulations take place at the beginning of the trial, after stimulus presentation, towards the end of the trial, and just before onset of the aversive noise (US). Specifically, the projection of decoding accuracy at the source space (see Figure 6) showed a strong involvement of early visual cortex at the beginning of the trial (from 0.2 s). Furthermore, the last 200 milliseconds prior to the onset of the US present a spatial shift in the discrimination between fear relevant or irrelevant cues by recruiting premotor and dorsal regions of the frontal lobe. These occipital and frontal areas implicated in the fear decoding revealed the same latencies and topography for all the three pairwise CS+ vs GS, indicating a similar cognitive processing during the task.

The time-GAL connectivity pattern showed the differences between the three GS responses. Higher direct generalization (black lines in Figure 6) is present in frontal areas of the CS+ vs. GS1 pairwise compared to the inverse generalization (red lines), while the CS+ vs. GS3 pairwise showed the contrary pattern with less direct generalization and a massive frontal involvement in inverse generalization. This frontal modulation was already established, but this final time-GAL step showed that it is most pronounced at the beginning and end of the trial. On the other hand, higher generalization in both direct and inverse direction of the visual and posterior tiers was found for the CS+ vs. GS3 compared with the CS+ vs. GS1 pairwise, indicating the contribution of this regions to discriminate better between different orientations. This posterior differences in generalization persisted throughout the entire trial. Taken together, results of the generalization connectivity matrices showed that the decoding itself (i.e. diagonal of the matrix) did not differ between the three comparisons made here, but the generalization across locations did.

As a control, we repeated the same analysis in the habituation phase of the fear conditioning paradigm. This phase is characterized by the absence of the aversive white noise,

i.e. the CS+ orientation is not paired with the US. Therefore, no threat is associated with any of the oriented Gabor patches.

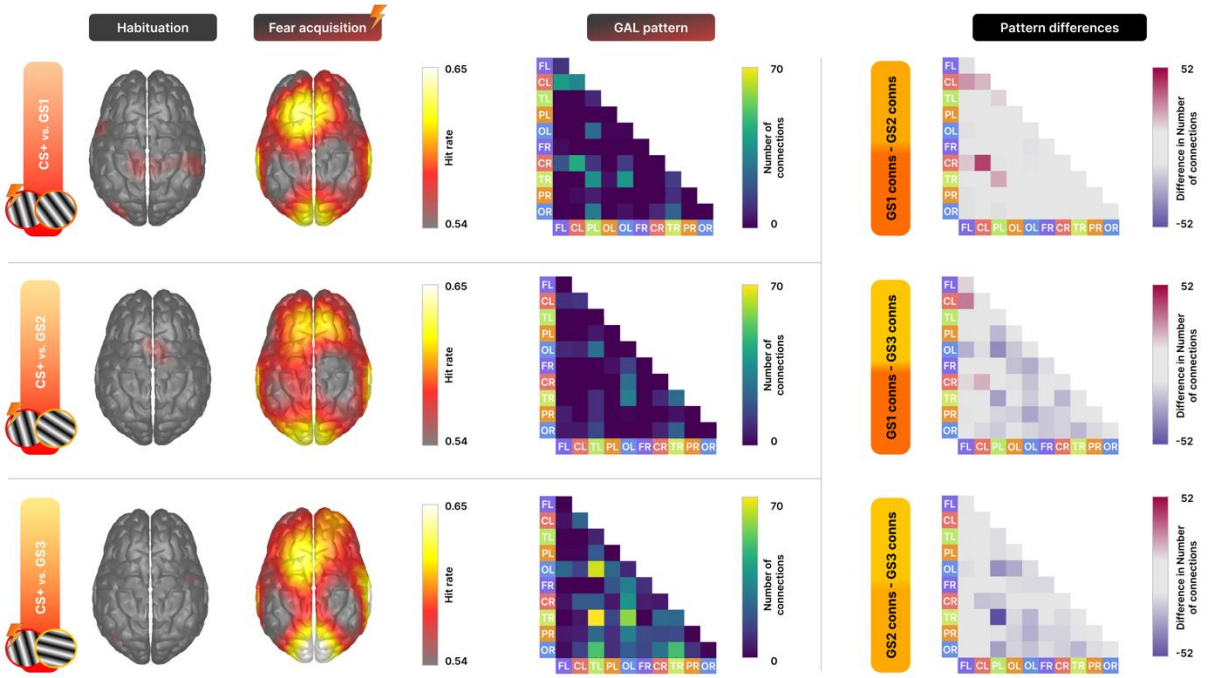

Figure 7. Changes in decoding and GAL connectivity patterns during fear acquisition. **Left**: Brain projection of decoding accuracy between CS+ and GS stimuli. The left brain refers to the habituation phase where CS+ stimulus was not yet paired with the aversive US, while the right brain corresponds with the fear acquisition phase where it acquired fear-related properties. **Middle**: GAL connectivity pattern for the three CS+ vs. GS pairwise based on the number of statistically significant generalization connections (Bonferroni-corrected $\alpha = 0.01$) within brain areas during the fear acquisition phase. **Right**: Differences in generalization connectivity pattern between the GAL from each one of the three pairwise. Abbreviations stand for F = frontal, C = central, T = temporal, P = parietal, O = occipital and L or R for left and right, respectively.

Interestingly, decoding-based GAL analysis conducted on the habituation phase did not reveal any statistically significant decoding (diagonal) or connection (triangles) from the entire

generalization connectivity matrix (p > Bonferroni-corrected 0.01). This lack of discriminability between oriented patches highlights the specificity of the effects observed during acquisition.

An alternative way of quantifying and illustrating time-GAL information is shown in Figure 7. Time-GAL information is readily reduced by organizing the density of connections between regions as presented in the middle column of Figure 7. The resulting matrices show the total amount of connections, both direct and inverse, between two areas. The right column of Figure 7 shows differences involving higher connectivity (red squares) in channels of central areas when comparing CS+ vs GS1, compared to other comparisons, while CS+ vs. GS3 showed more connections (blue squares) in temporal and posterior areas.

Discussion

The emergence of multivariate decoding techniques has made it possible to quantify and discriminate dynamic changes in neural patterns. Here, we present a decoding approach that uses the temporal patterns of neural population signals. The time-based decoder is then used to quantify the similarity of time courses across different locations. Furthermore, the combination of this decoding approach with a forward model visualizes the changing cross-decoding over time and thus illustrates the temporal sequence of spatial activation patterns.

The present study demonstrates the usage of the time-GAL method, employing two EEG datasets recorded during two tasks, both involving vision and emotion. In the first study, consistent with a large body of evidence (e.g., Liu et al., 2012), we found strong decoding during the period of the late positive potential (LPP) ranging from 0.5 s to 1 s post-stimulus. However, previous univariate studies rarely observed differences between pleasant and unpleasant content, supporting the notion that the multivariate time-GAL approach is more sensitive than univariate approaches. In addition, while univariate models (Friston et al., 1994) have occasionally identified condition differences between pleasant and unpleasant content during the LPP time window, the

MVPA-based approach extracted the spatial pattern differences and highlighted brain areas with shared decodable information. Therefore, the time-GAL has substantial potential for defining the spatio-temporal dynamics specific to processing emotional scene content.

Application of time-GAL to the aversive conditioning dataset (data set 2) illustrated how the method discriminated between neural responses to threat versus safety, after conditioning. Prior to conditioning, no decoding was found for the same stimulus pairs, highlighting the sensitivity of the method to acquired stimulus properties such as conditioned threat (Lonsdorf et al., 2017). In our example, dense anterior to posterior generalization connectivity patterns were found for the comparison of the fearful CS+ and the most similar GS condition, consistent with notions of sharpened orientation tuning as a function of experience.

The decoding methodology proposed in the present work aims to offer a new tool for the analysis and study of neurophysiological data. This approach leverages the rich temporal information contained in electrophysiological time series. Instead of using spatial information as traditionally carried out in fMRI (Haxby, 2012; Peelen & Downing, 2023) and M/EEG (Ashton et al., 2022; Bae & Luck, 2018, 2019), here we utilize the temporal dimension as features to feed the decoding model. To our knowledge, MVPA-based decoding approaches have not yet been applied to time-based features. Therefore, the present tool may be the first step towards suitable methods for characterization neural activity patterns while taking advantage of their high temporal resolution.

Decoding approaches have been applied to neuroimaging data in a diverse range of procedures. Bae and Luck (2018) have conducted time-resolved MVPA approaches using averages of trials in order to obtain better noise-ratio information and reduce computational resources. This approach obtains the evoked information of the trials, yet loses the induced information of each individual trial (David et al., 2006). Following their approach, we computed a similar version of the GAL procedure making use of averaged groups of trials (see Supplementary

1). Overall similar information was found in the generalization connectivity matrices using averaged data compared with the single-trial based method proposed in this paper. Interestingly, no statistically significant inverse generalization connections were found when using trial averages, potentially indicating the usage of non-phase locked activity for single-trial based decoding.

The time-GAL approach can be easily applied using the toolbox available in https://github.com/csea-lab/time-GAL. No statistical assumptions are required to be met and both stationary and non-stationary signals can be used for decoding. Therefore, this procedure can be utilized to analyze a wide range of cognitive and affective neurophysiological studies without any special preparation or restriction. Here, a current source density transformation was applied to more specifically examine spatial aspects of the EEG topography and to minimize effects of volume conduction.

The second step of the time GAL procedure offers information about how decodable information is generalized across the brain. The generalization connectivity maps obtained through the algorithm may be suitable to examine the configuration of dynamic cortical networks (Pessoa, 2014). Traditionally, these networks have been studied conducting functional connectivity analysis using measures that reflect the statistical dependencies between two time series, such as the Pearson correlation coefficient or other metrics derived from general linear models (Friston et al., 1994). This methodology is often applied to resting-state recordings (López-Sanz et al., 2017) and tends to involve filtering of the data to highlight a given aspect of the neural time series. Here we extract decodable information from the waveform itself at the single trial level, using all available data.

In conclusion, time-GAL is a novel algorithm for examining neural dynamics based on decoding. At its core, it is non-parametric in nature because it is based on decoding accuracy measured by leave-one-out cross-validation.

References

Ashton, K., Zinszer, B. D., Cichy, R. M., Nelson, C. A., Aslin, R. N., & Bayet, L. (2022). Time-resolved multivariate pattern analysis of infant EEG data: A practical tutorial. *Developmental Cognitive Neuroscience*, *54*, 101094. https://doi.org/10.1016/j.dcn.2022.101094

Bae, G.-Y., & Luck, S. J. (2018). Dissociable Decoding of Spatial Attention and Working Memory from EEG Oscillations and Sustained Potentials. *The Journal of Neuroscience*, *38*(2), 409–422. https://doi.org/10.1523/JNEUROSCI.2860-17.2017

Bae, G.-Y., & Luck, S. J. (2019). Reactivation of Previous Experiences in a Working Memory Task. *Psychological Science*, *30*(4), 587–595. https://doi.org/10.1177/0956797619830398

Bo, K., Cui, L., Yin, S., Hu, Z., Hong, X., Kim, S., Keil, A., & Ding, M. (2022). Decoding the temporal dynamics of affective scene processing. *NeuroImage*, *261*, 119532. https://doi.org/10.1016/j.neuroimage.2022.119532

Carrasco, C. D., Bahle, B., Simmons, A. M., & Luck, S. J. (2024). Using multivariate pattern analysis to increase effect sizes for event-related potential analyses. *Psychophysiology*, e14570. https://doi.org/10.1111/psyp.14570

David, O., Kilner, J. M., & Friston, K. J. (2006). Mechanisms of evoked and induced responses in MEG/EEG. *NeuroImage*, *31*(4), 1580–1591. https://doi.org/10.1016/j.neuroimage.2006.02.034

Di Russo, F., Martínez, A., Sereno, M. I., Pitzalis, S., & Hillyard, S. A. (2002). Cortical sources of the early components of the visual evoked potential: Cortical Sources of VEP. *Human Brain Mapping*, *15*(2), 95–111. https://doi.org/10.1002/hbm.10010

Friston, K. J., Jezzard, P., & Turner, R. (1994). Analysis of functional MRI time-series. *Human Brain Mapping*, *1*(2), 153–171. https://doi.org/10.1002/hbm.460010207

Gibney, K. D., Kypriotakis, G., Cinciripini, P. M., Robinson, J. D., Minnix, J. A., & Versace, F. (2020). Estimating statistical power for event-related potential studies using the late positive potential. *Psychophysiology*, *57*(2), e13482. https://doi.org/10.1111/psyp.13482

Grootswagers, T., Wardle, S. G., & Carlson, T. A. (2017). Decoding Dynamic Brain Patterns from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging Data. *Journal of Cognitive Neuroscience*, *29*(4), 677–697. https://doi.org/10.1162/jocn_a_01068

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., & Bießmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage*, *87*, 96–110. https://doi.org/10.1016/j.neuroimage.2013.10.067

Haxby, J. V. (2012). Multivariate pattern analysis of fMRI: The early beginnings. *NeuroImage*, *62*(2), 852–855. https://doi.org/10.1016/j.neuroimage.2012.03.016

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. *Science*, *293*(5539), 2425–2430. https://doi.org/10.1126/science.1063736

Junghofer, M., Elbert, T., Leiderer, P., Berg, P., & Rockstroh, B. (1997). Mapping EEG-potentials on the surface of the brain: A strategy for uncovering cortical sources. *Brain Topography*, *9*(3), 203–217. https://doi.org/10.1007/BF01190389

Junghofer, M., Elbert, T., Tucker, D. M., & Rockstroh, B. (2000). Statistical control of artifacts in dense array EEG/MEG studies. *Psychophysiology*, *37*(4), 523–532. https://doi.org/10.1111/1469-8986.3740523

Keil, A., Bernat, E. M., Cohen, M. X., Ding, M., Fabiani, M., Gratton, G., Kappenman, E. S., Maris, E., Mathewson, K. E., Ward, R. T., & Weisz, N. (2022). Recommendations and

publication guidelines for studies using frequency domain and time-frequency domain analyses of neural time series. *Psychophysiology*, *59*(5). https://doi.org/10.1111/psyp.14052

King, J.-R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: The temporal generalization method. *Trends in Cognitive Sciences*, *18*(4), 203–210. https://doi.org/10.1016/j.tics.2014.01.002

Lang, Bradley, & Cuthbert. (2008). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual.* (Gainesville, FL: University of Florida, NIMH Center for the Sutdy of Emotion and Attention).

Liu, Y., Huang, H., McGinnis-Deweese, M., Keil, A., & Ding, M. (2012). Neural Substrate of the Late Positive Potential in Emotional Processing. *The Journal of Neuroscience*, *32*(42), 14563–14572. https://doi.org/10.1523/JNEUROSCI.3109-12.2012

Lonsdorf, T. B., Menz, M. M., Andreatta, M., Fullana, M. A., Golkar, A., Haaker, J., Heitland, I., Hermann, A., Kuhn, M., Kruse, O., Meir Drexler, S., Meulders, A., Nees, F., Pittig, A., Richter, J., Römer, S., Shiban, Y., Schmitz, A., Straube, B., … Merz, C. J. (2017). Don't fear 'fear conditioning': Methodological considerations for the design and analysis of studies on human fear acquisition, extinction, and return of fear. *Neuroscience & Biobehavioral Reviews*, *77*, 247–285. https://doi.org/10.1016/j.neubiorev.2017.02.026

López-Sanz, D., Bruña, R., Garcés, P., Martín-Buro, M. C., Walter, S., Delgado, M. L., Montenegro, M., López Higes, R., Marcos, A., & Maestú, F. (2017). Functional Connectivity Disruption in Subjective Cognitive Decline and Mild Cognitive Impairment: A Common Pattern of Alterations. *Frontiers in Aging Neuroscience*, *9*, 109. https://doi.org/10.3389/fnagi.2017.00109

Mahmoudi, A., Takerkart, S., Regragui, F., Boussaoud, D., & Brovelli, A. (2012). Multivoxel Pattern Analysis for fMRI Data: A Review. *Computational and Mathematical Methods in Medicine*, *2012*, 1–14. https://doi.org/10.1155/2012/961257

Mandelkow, H., De Zwart, J. A., & Duyn, J. H. (2016). Linear Discriminant Analysis Achieves High Classification Accuracy for the BOLD fMRI Response to Naturalistic Movie Stimuli. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00128

Norcia, A. M., Appelbaum, L. G., Ales, J. M., Cottereau, B. R., & Rossion, B. (2015). The steady-state visual evoked potential in vision research: A review. *Journal of Vision*, *15*(6), 4. https://doi.org/10.1167/15.6.4

Nunez, P. L., Srinivasan, R., Westdorp, A. F., Wijesinghe, R. S., Tucker, D. M., Silberstein, R. B., & Cadusch, P. J. (1997). EEG coherency: I: statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales. *Electroencephalography and Clinical Neurophysiology*, *103*(5), 499–515. https://doi.org/10.1016/S0013-4694(97)00066-7

Peelen, M. V., & Downing, P. E. (2023). Testing cognitive theories with multivariate pattern analysis of neuroimaging data. *Nature Human Behaviour*, *7*(9), 1430–1441. https://doi.org/10.1038/s41562-023-01680-z

Pessoa, L. (2014). Understanding brain networks and brain organization. *Physics of Life Reviews*, *11*(3), 400–435. https://doi.org/10.1016/j.plrev.2014.03.005

Peyk, P., De Cesarei, A., & Junghöfer, M. (2011). ElectroMagnetoEncephalography Software: Overview and Integration with Other EEG/MEG Toolboxes. *Computational Intelligence and Neuroscience*, *2011*, 1–10. https://doi.org/10.1155/2011/861705

Rivolta, D., Woolgar, A., Palermo, R., Butko, M., Schmalzl, L., & Williams, M. A. (2014). Multi-voxel pattern analysis (MVPA) reveals abnormal fMRI activity in both the â€œcoreâ€ and â€œextendedâ€ face network in congenital prosopagnosia. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00925

Santos-Mayo, A., de Echegaray, J., & Moratti, S. (2022). Conditioned up and down modulations of short latency gamma band oscillations in visual cortex during fear learning in humans. *Scientific Reports*, *12*(1), 2652. https://doi.org/10.1038/s41598-022-06596-8

Wieser, M. J., Miskovic, V., & Keil, A. (2016). Steady-state visual evoked potentials as a research tool in social affective neuroscience. *Psychophysiology*, *53*(12), 1763–1775. https://doi.org/10.1111/psyp.12768

Supplementary 1

Traditionally, the MVPA-based decoding approach has been conducted using spatial information to classify between conditions (Haxby, 2012). Usually, information from the experimental trials is condensed in averaged groups of trials to reduce the computational resources needed to train the model (Bae & Luck, 2018, 2019). Therefore, classifier models requiring heavier computational resources, such as supported vector machine (SVM), can be used. However, in our work we employed linear discriminant analysis (LDA), a classifier method requiring less resources, thus allowing us to compute the generalization across location (GAL) methodology at the single-trial level.

Analysis of brain electrophysiological data following an evoked (averaging data) or induced (using single-trials) approach can derive different results (David et al., 2006; Santos-Mayo et al., 2022). For this reason, here we aimed to investigate whether the results of the GAL methodology using an evoked one similar as performed by Bae and Luck (2018) could differ from the induced approach developed in the main text. Particularly, in this complementary analysis we randomly divided the trials for each condition in three averaged groups using two for training the model and one for the test phase. Again, a leave one subject out cross validation was carried out training the model using *n*-1 subject and testing in the remaining subject. For simplicity and better comparison, ERP signal from the affective EEG task was employed since it benefits better from average than SSVEP signal.

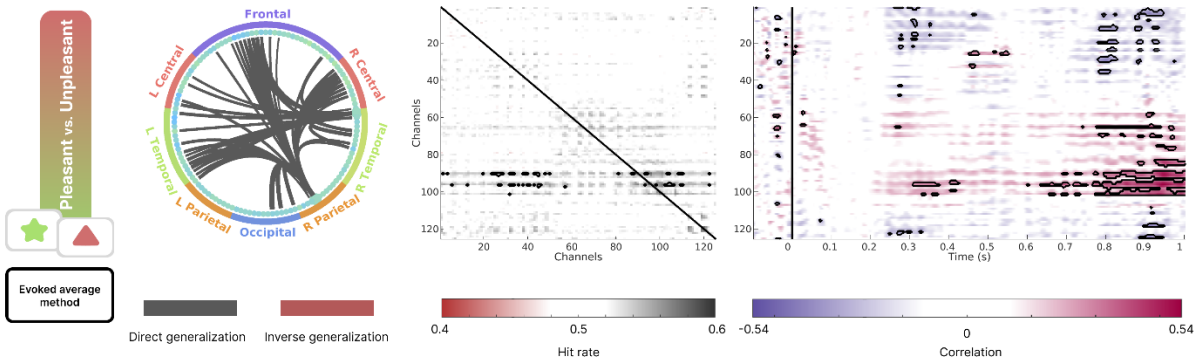

Figure S1. Generalization connectivity pattern and temporal distribution for affective pictures using an average-based evoked approach. **Left**: Circular graph showing the statistically significant generalization connections between 125 CSD channels. Black lines refer to those connections where both channels show similar decoding temporal patterns. **Center**: Generalization connectivity matrix showing decoding results (accuracy). Y axis corresponds with the channel data employed to train the model, while X axis refers to the channel data used in the test phase. Black contours indicate direct generalization connections after a Bonferroni-corrected statistical (α = 0.01) comparison with chance level (0.5). **Right**: Temporal correlation between labels of condition and data showing the positive or negative weights of the decoding model. Black contours indicate those time points whose p-value is below 0.01 for each channel.

Interestingly, we can observe a lack of inverse generalization connectivity, indicating that the evoked method does not find statistically significant neural patterns of opposite response to the stimuli. In addition, induced approach found more contribution of the electrode's information to the discrimination of the conditions (Figure 3 right) than here the evoked analysis (Figure S1 right) from 300 ms to 500 ms. These differences may highlight the capability of the MVPA tools to extract the information from single-trials contained at the induced level and lost when information is averaged. However, late cognitive processing from 800 ms to 1 s shows similar contribution patterns in similar electrodes indicating that process could be more stable and less oscillation-dependent. Thus, difference between averaging approaches may not differ in late latencies. Regarding the generalization connectivity patterns (Figure 3 left and Figure S1 left),

both analysis shows slight differences but preserving the same anterior-posterior generalization connectivity patterns.

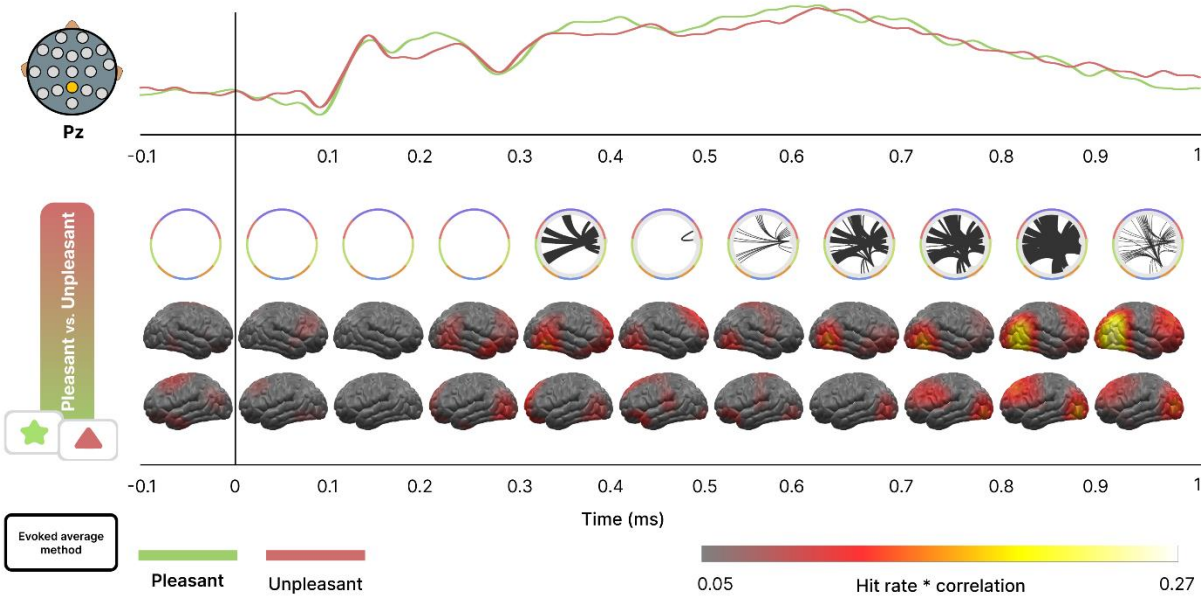

Figure S2. Time course of emotional content decoding and GAL connectivity patterns in affective pictures using an average-based evoked approach. **Top**: Electrophysiological ERP response to the presentation of different emotional content obtained at the parietal central (Pz) electrode for the three conditions (pleasant, neutral and unpleasant). **Middle**: Circular graphs show the GAL connectivity pattern at each time step for each pairwise condition. Brain's projections indicate the convolution of decoding rate by time correlation describing the fluctuation of the emotional process during the trial over the cortex.

Inspecting the decoding topographies (Figure S2 bottom), we can observe similar decoding patterns than those obtained using an induced methodology (Figure 4 bottom). Specifically, same anterior and posterior distribution of the neural patterns are found, overall, during the last part of the trial (800 ms to 1 s). This similarity shows that same areas and latencies are involved in the discrimination between pleasant and unpleasant conditions even when an induced or evoked methodologies are conducted, thus suggesting that both approaches are valid to analyze and extract conclusions from data.

In addition, GAL connectivity patterns shows their activity in the same latencies as previously indicated (Figure S2 top). In contrast with the induced methodology, less GAL activity can be found during the first part of the trial (300 ms to 500 ms) using the evoked approach. These results could highlight the communality of cognitive processes in late latencies discoverable from both induced and evoked approaches. In contrast, the induced methodology shows optimal to extract the information related to the oscillatory bands phase during the early latencies.