A Appendix

A.1 Supplementary methods

A.1.1 ICA data preprocessing

For each subject, we performed an ICA decomposition on his/her EEG recorded brain potentials (Makeig et al., 1996). Briefly, if $\mathbf{x} \in \mathbb{R}^N$ is an N-dimensional random vector representing EEG in N channels, we estimated a mixing matrix $A \in \mathbb{R}^{N \times N}$ so that

$$\mathbf{x} = A\mathbf{s} \tag{A.1}$$

and the components of the random vector $\mathbf{s} \in \mathbb{R}^N$ were maximally independent. These components are called the independent components (ICs) through the manuscript. We used the AMICA algorithm (Palmer et al., 2007) with one model to compute this decomposition. After ICA decomposition, ICs with ECD location outside the brain were automatically removed, and non-brain ICs (e.g., components corresponding to eye blinks, lateral eye movements, muscle activations, and bad channels) were manually removed. We kept, an average, of 26 components per subject, out of the $N = 33$ components obtained in the ICA decomposition.

A.1.2 Equivalent current dipole location

The scalp map of the ith IC is the ith column of the mixing matrix A in (A.1). For each IC, the location of the electric current dipole whose projection in the scalp best matched the IC scalp map, i.e., the equivalent current dipole (ECD) location, was estimated using the DIPFIT2 plugin of the EEGLAB software. The location of this dipole is the ECD location of the IC.

A.1.3 Clustering of ICs

All ICs of all subjects were grouped into clusters according to the proximity of their ECD locations (see Section A.1.2) using the k-means algorithm. This algorithm groups points of a D-dimensional metric space into clusters, in such a way that the sum of the distances between points and their corresponding cluster centroid is minimized. We represented each IC by the three-dimensional point given by the Talairach coordinates of its ECD location. A free parameter in k-means is the number of clusters. We set this parameter to 17, in order to obtain clusters of reasonable coarseness. Clusters 8, 12, and 16 were not analyzed because they contained too few ICs, 4, 4 and 5, respectively (Table A.1).

A.1.4 Circular statistics concepts

This section introduces concepts from circular statistics (Fisher, 1996) used to define ITC and DMP. Given a set of circular variables (e.g., phases), $\theta_1, \ldots, \theta_N$, we associate to each circular variable a two-dimensional unit vector. Using notation from complex numbers, the unit vector associated with variable θ_i is:

$$vec(\theta_i) = e^{j\theta_i} \tag{A.2}$$

The *resultant vector*, $\mathbf{R}$, is the sum of the associated unit vectors:

$$\mathbf{R}(\theta_1, \ldots, \theta_N) = \sum_{i=1}^{N} vec(\theta_i) \tag{A.3}$$

The *mean resultant length*, $\bar{R}$, is the length of the resultant vector divided by the number of circular variables:

$$\bar{R}(\theta_1, \ldots, \theta_N) = \frac{1}{N}|\mathbf{R}(\theta_1, \ldots, \theta_N)| \tag{A.4}$$

The *circular variance*, CV, is one minus the mean resultant length:

$$CV(\theta_1, \ldots, \theta_N) = 1 - \bar{R}(\theta_1, \ldots, \theta_N) \tag{A.5}$$

The *mean direction*, $\bar{\theta}$, is the angle of the resultant vector:

$$\bar{\theta}(\theta_1, \ldots, \theta_N) = \arg(\mathbf{R}(\theta_1, \ldots, \theta_N)) \tag{A.6}$$

Note that the mean direction (and therefore the DMP, Section 2.5) is not defined when the resultant vector is zero, since the angle of the zero vector is undefined.

A.1.5 ITC and Peak ITC frequency

The Inter-Trial Coherence (ITC) is a measure of ITPC resulting from averaging phase information among multiple epochs (Tallon-Baudry et al., 1996; Delorme and Makeig, 2004). To compute ITC we extracted epochs from one second before to three seconds after the presentation of standards. Then, a continuous wavelet transform was performed on these epochs, using the Morlet wavelet with three significant cycles, eight octaves, and 12 frequencies per octave. With a 250 Hz EEG sampling rate, the previous parameters furnished a Morlet transform with a time resolution of 70.71 ms, and a frequency resolution of 4.5 Hz, both at 10 Hz. The function `cwt` in the `Rwave` package of the language `R` (R Core Team, 2012) was used to compute the continuous wavelet transform. This transform provided the phases of every trial, for frequencies between 0.5 and 128 Hz, and

for times between the start and end of the epoch. The ITC of a set of phases $\theta_1, \ldots, \theta_n$ (gray vectors in Figure 2) is the mean resultant length ($\bar{R}$, Eq. A.4) of these phases (length of black vector in Figure 2):

$$ITC(\theta_1, \ldots, \theta_N) = \bar{R}(\theta_1, \ldots, \theta_N) \tag{A.7}$$

For each IC of every subject, standard modality, and attended modality, we used the corresponding set of epochs to calculate ITC between the start and end times of the epochs, and between 0.5 Hz to 125 Hz. We selected the peak ITC between 100 and 500 ms after the presentation of the standard at time zero, and between 1 and 14 Hz. The frequency corresponding to this peak (i.e., the peak ITC frequency) was then used to measure the single-trial phase, as described in Section A.1.6. The median peak ITC frequency and its 95% confidence interval were 4.93 Hz and [4.59, 5.21] Hz, respectively, and the median time of the ITC peak and its 95% confidence interval were 215 ms and [212, 226] ms, respectively. ITC between -200 and 500 ms around the presentation of attended visual standards, from IC 5 of subject av130a are shown in Figure A.1. The black cross in this figure marks the peak ITC value, occurring at peak frequency 8.15 Hz.

A.1.6 Measuring phases in single trials

After selecting the peak ITC frequency (Section A.1.5) for an IC of a subject, a standard modality, and an attended modality, we computed a Gabor transform of all epochs at this frequency. These epochs started one second before and ended three seconds after the presentation of standards. For this calculation, we used the function `cgt` in the `Rwave` package of the language `R` (R Core Team, 2012). We adjusted the scale of the Gabor's Gaussian window to obtain three significant sinusoids at the peak ITC frequency. The phases of single trials were extracted

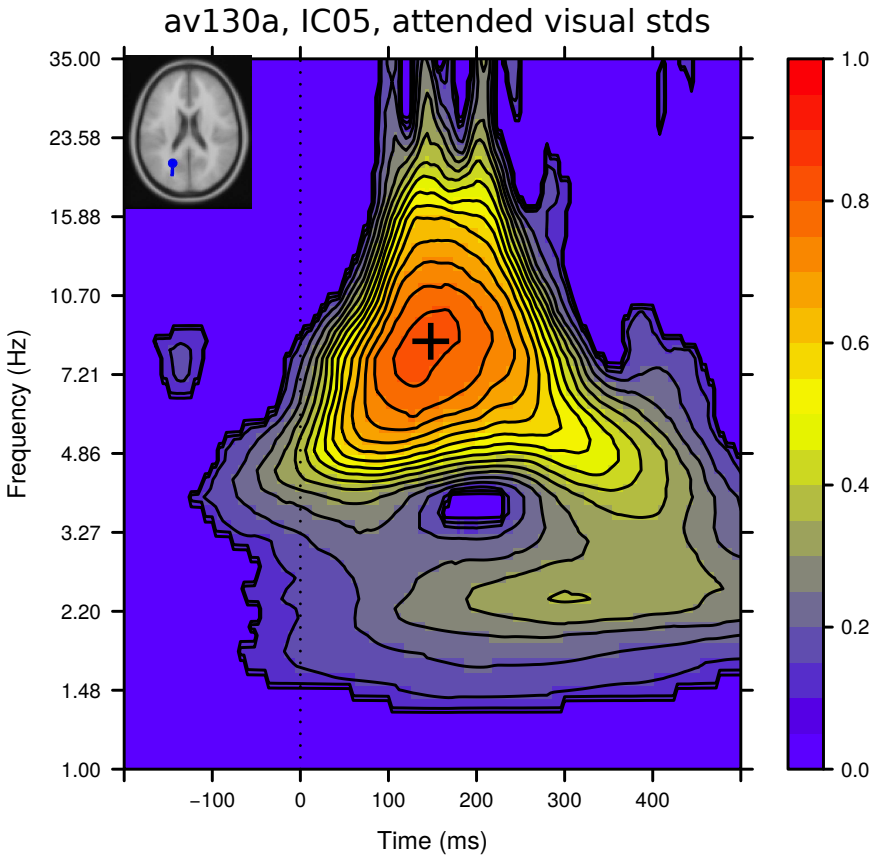


Figure A.1: Example set of ITC values and selected peak from IC 5 of subject av130a and attended visual standards. The peak ITC value (black cross) was selected between 100 and 500 ms and between 1 and 14 Hz. The selected peak ITC frequency, time, and value were 8.52 Hz, 148 ms, and 0.83, respectively. Non-significant ITC ($p>0.01$, Rayleigh uniformity test) is masked in light blue.

from the complex coefficients of this Gabor transform.

A.1.7 Relation between the averaged DMP and the ITC

The DMP is a measure of phase decoherence, while the ITC is a measure of phase coherence. Here we derive an upper bound for the averaged DMP, $\overline{DMP}$, by ITC (Proposition 1). From this bound we infer that to minimal $\overline{DMP}$ (i.e., $\overline{DMP} \simeq 0$) correspond large ITC (i.e., $ITC \simeq 1$), and that the maximal $\overline{DMP}$ is 1/2.

Proposition 1. $\overline{DMP} \leq \frac{1}{2} - \frac{ITC}{2}$

Proof. We first rewrite DMP

$$\begin{aligned} DMP(\theta_i|\{\theta_1,\ldots,\theta_N\}) &= CV(\theta_i, \bar{\theta}(\theta_1,\ldots,\theta_N)) \\ &= 1 - \bar{R}(\theta_i, \bar{\theta}(\theta_1,\ldots,\theta_N)) \\ &= 1 - \frac{1}{2}|\mathbf{R}(\theta_i, \bar{\theta}(\theta_1,\ldots,\theta_N))| \\ &= 1 - \frac{1}{2}|vec(\theta_i) + vec(\bar{\theta}(\theta_1,\ldots,\theta_N))| \end{aligned} \quad (A.8)$$

The first equality comes from Eq. 1, the second one from Eq. A.5, the third one from Eq. A.4, and the fourth one from Eq. A.3. We also rewrite $vec(\bar{\theta}(\theta_1,\ldots,\theta_N))$:

$$\begin{aligned} vec(\bar{\theta}(\theta_1,\ldots,\theta_N)) &= vec(\arg(\mathbf{R}(\theta_1,\ldots,\theta_N))) \\ &= \frac{\mathbf{R}(\theta_1,\ldots,\theta_N)}{|\mathbf{R}(\theta_1,\ldots,\theta_N)|} \\ &= \frac{\mathbf{R}(\theta_1,\ldots,\theta_N)}{N\bar{R}(\theta_1,\ldots,\theta_N)|} \end{aligned} \quad (A.9)$$

The first equality comes from Eq. A.6, the second one from the fact that *vec* applied to arg of any vector gives the vector normalized to unit length, the third equality comes from Eq. A.4. Then,

$$
\begin{aligned}
1-\overline{DMP} &= 1-\frac{1}{N}\Sigma_{i=1}^{N} DMP(\theta_i|\{\theta_1,\ldots,\theta_N\}) \\
&= \frac{1}{2N}\Sigma_{i=1}^{N}|vec(\theta_i)+vec(\bar{\theta}(\theta_1,\ldots,\theta_N))| \\
&\geq \frac{1}{2N}\left|\Sigma_{i=1}^{N} vec(\theta_i)+N\,vec(\bar{\theta}(\theta_1,\ldots,\theta_N))\right| \\
&= \frac{1}{2N}\left|\mathbf{R}(\theta_1,\ldots,\theta_N)+\frac{\mathbf{R}(\theta_1,\ldots,\theta_N))}{\bar{R}(\theta_1,\ldots,\theta_N)}\right| \\
&= \frac{1}{2N}\left(1+\frac{1}{\bar{R}(\theta_1,\ldots,\theta_N)}\right)|\mathbf{R}(\theta_1,\ldots,\theta_N)| \\
&= \frac{1}{2}\left(1+\frac{1}{\bar{R}(\theta_1,\ldots,\theta_N)}\right)\bar{R}(\theta_1,\ldots,\theta_N) \\
&= \frac{1}{2}\left(\bar{R}(\theta_1,\ldots,\theta_N)+1\right) \\
&= \frac{ITC}{2}+\frac{1}{2}
\end{aligned}
\tag{A.10}
$$

The first equality comes from the definition of $\overline{DMP}$, the second one from Eq. A.8, the third one from the triangle inequality, the fourth one from Eq. A.3 and Eq. A.9, the fifth one from taking $\mathbf{R}(\theta_{\mathbf{1}},\ldots,\theta_{\mathbf{N}})$ as common factor, the sixth one from Eq. A.4, the seventh one from distributing $\bar{R}(\theta_1,\ldots,\theta_N)$, and the eight one from Eq. A.7. The proposition follows by re-arranging Eq. A.10. □

Because $0 \leq DMP \leq 1$ it follows that $0 \leq \overline{DMP} \leq 1$. When ITC is maximal (i.e., ITC=1) Proposition 1 shows that $\overline{DMP} \leq 0$, thus $\overline{DMP}$ must achieve it minimum (i.e., $\overline{DMP} = 0$). Also, this proposition shows that 1/2 is an upper bound for $\overline{DMP}$, corresponding to zero ITC.

A.1.8 Equivalence between the minimization of the ridge regression error function and the maximization of the log posterior of the model coefficients

Proposition 2. *Given the likelihood function in Eq. 5, and the priors in Eq. 6 and Eq. 7, the coefficients* $\mathbf{w}$ *maximizing the logarithm of the posterior distribution*

$J(\mathbf{w})$ in Eq. 8 minimize the ridge-regression error function RMSE$(\mathbf{w})$ in Eq. 4.

Proof. We first rewrite the joint pdf $P(\mathbf{y}, \mathbf{w}, \tau, \alpha, \Phi)$ as:

$$
\begin{aligned}
P(\mathbf{y}, \mathbf{w}, \tau, \alpha, \Phi) &= P(\mathbf{y}|\mathbf{w}, \tau, \Phi) P(\mathbf{w}, \tau|\alpha) P(\alpha) \\
&= N(\mathbf{y}|\Phi\mathbf{w}, \tau^{-1} I) N(\mathbf{w}|\mathbf{0}, (\tau\alpha)^{-1} I) \mathrm{Gam}(\tau|a_0, b_0) \mathrm{Gam}(\alpha|c_0, d_0) \\
&= \prod_{n=1}^{N} N(y[n]|\langle \mathbf{x}[n, \cdot], \mathbf{w} \rangle, \tau^{-1}) \\
&\quad \prod_{k=1}^{K} N(w[k]|0, (\tau\alpha)^{-1}) \\
&\quad \mathrm{Gam}(a_0, b_0) \mathrm{Gam}(c_0, d_0) \\
&= \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\tau^{-1}}} \exp\left(-\frac{(y[n] - \langle \mathbf{x}[n, \cdot], \mathbf{w} \rangle)^2}{2\tau^{-1}} \right) \\
&\quad \prod_{k=1}^{K} \frac{1}{\sqrt{2\pi(\tau\alpha)^{-1}}} \exp\left(-\frac{w[k]^2}{2(\tau\alpha)^{-1}} \right) \\
&\quad \mathrm{Gam}(\tau|a_0, b_0) \mathrm{Gam}(\alpha|c_0, d_0)
\end{aligned} \tag{A.11}
$$

The first equality comes from applying Bayes rule, the second one from substituting Eq. 5, Eq. 6, and Eq. 7 into the first equality, the third one from the fact that the Gaussian distributions in the second equality are independent, and the last one from the definition of a Gaussian distribution.

$$
\begin{aligned}
\arg\max_{\mathbf{w}} J(\mathbf{w}) &= \arg\max_{\mathbf{w}} \log P(\mathbf{w}|\mathbf{y},\tau,\alpha,\Phi) \\
&= \arg\max_{\mathbf{w}} \log P(\mathbf{y},\mathbf{w},\tau,\alpha,\Phi) \\
&= \arg\max_{\mathbf{w}} \left[\sum_{n=1}^{N} -\frac{(y[n]-\langle \mathbf{x}[n,\cdot],\mathbf{w}]\rangle)^2}{2\tau^{-1}} + \sum_{k=1}^{K} -\frac{w[k]^2}{2(\tau\alpha)^{-1}} \right] \\
&= \arg\min_{\mathbf{w}} \left[\sum_{n=1}^{N} \frac{(y[n]-\langle \mathbf{x}[n,\cdot],\mathbf{w}]\rangle)^2}{2\tau^{-1}} + \sum_{k=1}^{K} \frac{w[k]^2}{2(\tau\alpha)^{-1}} \right] \\
&= \arg\min_{\mathbf{w}} \left[\sum_{n=1}^{N} (y[n]-\langle \mathbf{x}[n,\cdot],\mathbf{w}]\rangle)^2 + \sum_{k=1}^{K} \frac{w[k]^2}{\alpha^{-1}} \right] \\
&= \arg\min_{\mathbf{w}} \left[\mathrm{MSE}(\mathbf{w}) + \alpha||\mathbf{w}||^2 \right] \\
&= \arg\min_{\mathbf{w}} \mathrm{RMSE}(\mathbf{w})
\end{aligned}
$$

The first equality comes from the definition of $J(\mathbf{w})$ in Eq. 8, the second one from the fact that, by Bayes rule, $P(\mathbf{w}|\mathbf{y},\tau,\alpha,\Phi)$ equals $P(\mathbf{y},\mathbf{w},\tau,\alpha,\Phi)$ times a factor independent of $\mathbf{w}$ and thus $\arg\max P(\mathbf{w}|\mathbf{y},\tau,\alpha,\Phi) = \arg\max P(\mathbf{y},\mathbf{w},\tau,\alpha,\Phi)$, the third one by discarding the terms not including $\mathbf{w}$ in Eq. A.11 and taking the logarithm, the fourth one from the fact that the maximum of a function is the minimum of the negative of that function, the fifth one from the fact that arg min of a function does not change by multiplying it by a constant, the sixth one from the definition of MSE($\mathbf{w}$) in Eq. 3, and the final one from the definition of RMSE($\mathbf{w}$) in Eq. 4. □

A.1.9 Adjustment for multiple comparisons in correlation tests

We followed procedures detailed in Westfall and Young (1993) and summarized next. For a family of hypothesis H_1 vs. H_1', H_2 vs. H_2', ..., H_k vs. H_k', we aim to control the Familywise Error Rate (FWE) defined as FWE=P(Reject at least one H_i|all H_i are true). For this we define adjusted P values, $\tilde{p}_i,\ i = 1,\ldots,k$, and we reject H_i at FWE=α if and only if $\tilde{p}_i \leq \alpha$. We denote observed values with lower-

case and the corresponding random variables with uppercase. The definition of the adjusted P value, $\tilde{p}_i$, should guarantee that FWE=α. That is, α=FWE=P(Reject at least one H_i at level α|all H_i are true)=P($\min_{1\leq j\leq k}\tilde{P}_j \leq \alpha$|all H_i are true). We define $\tilde{p}_i = P(\min_{1\leq j\leq k} P_j \leq p_i|$all H_i are true), where p_i is the observed non-adjusted P value. When the P values P_j are independent it can be shown (Westfall and Young, 1993, p. 47) that the previous definition leads to an exact multiple comparison method (i.e., FWE=α).

In our tests of correlation, the hypothesis H_i (H_i') states that the correlation coefficient between two sequences $\{a_1,\ldots,a_n\}$ and $\{b_1,\ldots,b_n\}$ corresponding to the ith cluster, standard modality, and attended modality is zero (different from zero). To compute the adjusted P value $\tilde{p}_i$ we use a resampling procedure. We first generate nResamples=5,000 samples of $\min_{1\leq j\leq k} P_j$ under the null hypothesis that all H_i are true and then estimate $\tilde{p}_i$ as the proportion of samples smaller or equal than the observed non-adjusted P value p_i. To generate a sample of $\min_{1\leq j\leq k} P_j$ under the null hypothesis, for each cluster, standard modality, and attended modality we: (1) obtain the sequences $\{a_1,\ldots,a_n\}$ and $\{b_1,\ldots,b_n\}$ to test for significant correlation (e.g., a_i="correlation coefficient between models' decodings and SFPDs for a subject i" and b_i="error rate for subject i" in Figure 5a), (2) shuffle the sequence $\{b_1,\ldots,b_n\}$ (to be under the null hypothesis), (3) compute the P value of the correlation coefficient between the $\{a_1,\ldots,a_n\}$ and the shuffled $\{b_1,\ldots,b_n\}$. The sample of $\min_{1\leq j\leq k} P_j$ under the null hypothesis is then the minimum of all P values generated in step (3) across all clusters, standard modalities, and attended modalities.

We adjusted for multiple comparison correlation tests between (a) models' decodings and SFPDs (Figure 3j-l, colored dots in Figure 4, Table A.2), and (b) models' decoding powers and subjects' behavioral measures (Figure 5, daggers in Figure 4, Tables A.3 and A.4). We begun by defining the family of hypothesis H_1 vs. H_1', ..., H_k vs. H_k' in the definition of the FWE, see above. In (a) we used 19

families of hypothesis, one per subject. The family for subject s contained hypothesis concerning correlation coefficients between model decodings and SFPDs, of all models for subject s across clusters, standard modalities, and attended modalities, that were significantly different from the intercept-only model ($p<0.01$; likelihood-ratio permutation test, Section A.1.12). The null ith hypothesis for subject s was H_i^s={the correlation coefficient between models' decodings and SFPDs for subject s and the ith combination of cluster, standard modality, and attended modality is zero}. The mean and standard deviation of the number of pairs of hypothesis in a family of hypothesis for a subject was 29.37 ± 12.64. In (b) we used one family of hypothesis for error rates and another one for mean reaction times. Each of these families contained hypothesis regarding the correlation between models' decoding powers and subjects' behavioral measures (error rates and mean reaction times) across all clusters, standard modalities, and attended modalities. Each of these correlations was evaluated across all models corresponding to a cluster, standard modality, and attended modality, as in Figure 5. The ith null hypothesis for behavioral measure b was H_i^b={the correlation coefficient between models' decoding accuracies and subjects' behavioral measure b for the ith combination of cluster, standard modality, and attended modality equals is zero}. Each family contained 56 pairs of hypothesis (14 clusters $\times$ 2 standard modalities $\times$ 2 attended modalities).

A.1.10 Calculation of robust correlation coefficients and corresponding P values

We used skipped measures of correlation (Wilcox, 2012) to characterize the association between pairs of variables in a way that was resistant to the presence of outliers. The calculation of these measures, and the estimation of their bootstrap confidence intervals, followed the procedures described in (Wilcox, 2012; Pernet et al., 2013).

Skipped correlations were obtained by checking for the presence of outliers, removing them, and applying some correlation coefficient to the remaining data (Wilcox, 2012, Chapter 9). We only considered bivariate outliers, which were selected using the function `outpro` freely available from Wilcox (2017).

To estimate the association between single-trial decoding errors and SFPDs (e.g., Figure 3j-l), we used the skipped Pearson correlation coefficient (i.e., after outlier removal, the remaining data points were correlated using the Pearson product moment correlation coefficient; function `pcor` from Wilcox (2017)). The use of this correlation coefficient requires that the marginals of the data are approximately normal, as it was the case in the previous data.

The averaged behavioral data tended to be bimodally distributed, with a group of subjects displaying better performance and another group showing worse performance. Thus, the marginals of the averaged behavioral data were not approximately normal, and it was not possible to use the skipped Pearson correlation coefficient to asses the association between average behavioral data and models' decodings. Instead, we quantified this association using the skipped Spearman correlation coefficient (i.e., after outlier removal, the remaining data points were correlated using the Spearman rank correlation coefficient; function `spear` from Wilcox (2017)). The use of this correlation coefficient does not assume that the distribution of the correlated variables is bivariate normal, but requires that their association be monotonic, as it was the case for our averaged data.

To compute P values corresponding to hypothesis tests of robust correlation coefficients we used permutation tests following the procedure given in Pernet et al. (2013). Given a dataset of two sequences to correlate, we (1) removed bivariate outliers from this dataset, (2) computed the correlation coefficient between the sequences in the outliers-removed dataset (using functions `pcor` and `spear` from Dr. Wilcox's website for Pearson and Spearman correlation coefficients, respectively), (3) computed nResamples=5000 correlation coefficients between the

first sequence and a random permutation of the second sequence in the outliers-removed dataset (again using functions `pcor` or `spear`), and (4) we calculated a two-sided P value as the proportion of correlation coefficients computed in step (3) whose absolute value was larger than that of the correlation coefficient computed in step (2).

A.1.11 Construction of ERP-superposition-removed datasets

For the switch-to-vision (LOOK) and the switch-to-audition (HEAR) warning signals of each dataset (i.e., of each subject, component, attended modality, and standard modality) we computed ERPs by calculating the mean of their ICA-component activations between one second before and three seconds after the presentation of the warning signal.

For each sample time we calculated the P value of a two-sided t-test checking whether the mean ICA-component activation at that time was different from zero. We corrected these P values for multiple comparison using the Bonferroni method.

For each warning signal ERP we defined the last-significant lag as the last ERP sample time with P value less than 0.05. For each dataset we built a corresponding ERP-superposition-removed dataset by removing standards following the preceding warning signal by less than the last-significant lag of the ERP of this warning signal.

A.1.12 Additional statistical information

Crossvalidated decodings All models decodings were cross validated using the leave-one-out method with the function `crossval` of the package `bootstrap` of `R` (R Core Team, 2012).

95% bootstrap CIs for regression coefficients (Figure 3m-o) We performed 2,000 ordinary bootstrap resamples of the trials (with the function `boot` of the package `boot` of `R` (R Core Team, 2012)), and for each resample we estimated a

set of regression coefficients, as described in Section 2.7. Having estimated 2,000 sets of regression coefficients, we computed 95% percentile confidence intervals with the function `boot.ci` of package `boot` of `R` (R Core Team, 2012).

95% bootstrap CI for difference in paired means of correlation coefficients obtained from the original versus a surrogate dataset (Figure 7). We performed 2,000 ordinary bootstrap resamples of the pairs of correlation coefficients obtained from the original versus a surrogate dataset (with the function `boot` of the package `boot` of `R` (R Core Team, 2012)). For each resample, and for each sample point, we subtracted the correlation coefficient of the original minus that of the surrogate dataset. Then we calculated the mean difference across all sample points. Having estimate 2,000 bootstrap means of pairwise differences, we computed a 95% percentile bootstrap confidence interval with the function `boot.ci` of package `boot` of `R` (R Core Team, 2012).

95% bootstrap CI for difference in averaged DMP between trials with the longest and shortest SFPD (Figures 3g-i). We calculated 2,000 stratified bootstrap resamples of the 20% of trials with the shortest and longest SFPDs, using the function `boot` of the package `boot` of `R` (R Core Team, 2012) with the `strata` option. For each resample, and for each sample point, we subtracted from the mean DMP across the 20% of trials with longest SFPD the mean DMP across the 20% of trials with shortest SFP. Having estimated 2,000 bootstrap differences in mean DMP between trials with the longest and shortest SFPD, we computed a 95% percentile bootstrap confidence interval for these differences with the function `boot.ci` of package `boot` of `R` (R Core Team, 2012).

Likelihood-ratio permutation test for linear model We used it to test the alternative hypothesis, H1, that there exists an association between the dependent and independent variables of a decoding model against the null one, H0, that no such association exists. The statistic for this test was the logarithm of the ratio between the likelihood of the data given a full model and that given a null model.

A full model associates all independent variables with the dependent one, while a null model associates only the constant term with the dependent variable. The likelihood of the data given a model is derived from Eq. 5. The test was conducted as follows. We first measured the test statistic in the original dataset, T_{obs}. Then, we built the distribution of the test statistic under H0, by generating R=2000 datasets where there was no association between the dependent and independent variables (by permuting the dependent value assigned to independent values) and measuring the test statistic in these datasets. The one-sided P value of the test is the proportion of samples of the test statistic under H0 that are larger than T_{obs}.
Normalized cross-correlation (Section 3.5) We normalized the cross-correlation (at lag zero) between two time series in such a way that it returned one when correlating a time series with itself. We defined:

$$ncor(x, y) = \frac{cor(x, y)}{\sqrt{cor(x, x)}\sqrt{cor(y, y)}}$$

where x and y are two time series, and $cor(x, y)$ is the unnormalized cross-correlation (at lag zero) between these time series. The normalized cross-correlation was computed with the function `ccf` of the package `stats` of `R` (R Core Team, 2012).

A.1.13 Selection of the optimal maximum SFPD

To fit decoding models we used data from standards that were presented before an optimal maximum SFPD after the warning signal. For models fitted to data from IC 5 of subject av130a, and unattended visual standards, Figure A.2 plots the decoding power of models as a function on their maximum SFPD. We see that the decoding power of models varied smoothly as a function of this maximum SFPD. For each estimated model we selected the maximum SFPD that optimized the

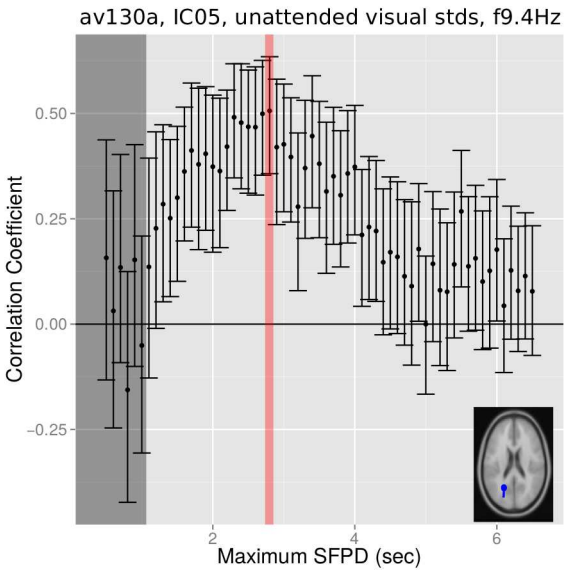


Figure A.2: Selection of the optimal maximum SFPD for the decoding model for IC 5 of subject av130a, unattended visual standards, and phase measured at 9.4 Hz. Decoding models were estimated with maximum SFPDs between 500 ms and the largest SFPD in a set of trials, in steps of 100 ms (abscissa), and the correlation coefficient were computed between models' decodings and experimental SFPDs (ordinate). The SFPD at which this correlation coefficient was largest was selected as the optimal maximum SFPD (red bar). To avoid possible modulations of ITPC by the warning signal, we excluded from this selection maximum SFPDs shorter than one second (gray bar).

decoding power of the model as the optimal maximum SFPD. We chose maximum SFPDs between one second and the largest SFPD of any trial in the data. We did not selected maximum SFPDs shorter than one second since ITPC following the presentation of standards close to the warning signal may be contaminated by evoked potentials by this warning signal. For IC 5 of subject av130a and unattended visual standards, the optimal maximum SFP duration was 2.8 seconds (time highlighted in red in Figure A.2).

A.2 Controls on the SFP effect on ITPC

From the successful decoding of SFPDs from ITPC evoked by standards we inferred that SFPD modulates ITPC evoked by standards. In Section A.2.1 we presented evidence supporting the inference that in the SFP effect on ITPC the modulated variable is the ITPC evoked by standards, and in Section A.2.2 we support the hypothesis that the modulating variable is the SFPD.

A.2.1 Modulations of ITPC used by decoding models do not seem to be generated by warning signals

The warning signal presented before a standard could have modulated the ITPC in the 500 ms-long time window following the presentation of a standard, and only this modulation could have allowed models to decode when the warning signal was presented prior to the standard at time zero (Figure 1c). In this alternative hypothesis, the occurrence of standards should be irrelevant and it should be possible to decode from any 500 ms-long time window starting at any time t_0 after the presentation of the warning signal that the warning signal was presented t_0 milliseconds before the start of the window.

We tested this alternative hypothesis by building surrogate datasets, identical to the original ones, with the only exception that epochs were aligned to pseudo-random times, instead of being aligned to the presentation time of standards (Section 2.4.1). For each original dataset, we generated 30 surrogate datasets. For each surrogate dataset, we fitted a decoding model (in exactly the same way as with the original dataset), and computed the correlation coefficient between model decodings and SFPDs. We used the median of these 30 correlation coefficients to measure the decoding power of models fitted to surrogate datasets. For models derived from ICs in the left parieto-occipital cluster 4 and the attended visual standards, Figure A.3 shows that the decoding power of models fitted to original

datasets was significantly larger than that of models fitted to corresponding surrogate datasets. The median pairwise difference between the decoding power of models fitted to original datasets minus that of models fitted to surrogate datasets was 0.21, which was significantly larger than zero (95% confidence interval [0.17, 0.25], Section A.1.12).

Note that the estimation of the decoding power for models fitted to surrogate datasets was computationally very expensive. For each cluster, standard modality, and attended modality, we fitted approximately 28,500 models, corresponding to 19 subjects $\times$ around 50 maximal SFPDs to select the optimal one (Section A.1.13) $\times$ 30 repetitions of the randomization procedure. For this reason, we limited this control to one representative cluster, standard modality, and attended modality.

This control furnishes evidence against the possibility that the observed modulations on ITPC were evoked by the warning signal, since modulations of ITPC by the warning signal should not be affected by the randomization of onsets of standards.

A.2.2 The SFP duration generates the SFP effect on ITPC

It is possible that from DMP evoked by standards one could reliably decode any variable, and not only the SFPD. This would be the case, for example, if a decoding model was overfitted to data. To control for this possibility, we proceed as in Section A.2.1, but with surrogate datasets with shuffled SFPDs. That is, instead of assigning to each trial its corresponding SFPD, we assigned the SFPD of a randomly chosen (without replacement) trial. Since the SFPD is the dependent variable in the decoding model (Section 2.6), the surrogate dataset only shuffled the dependent values of this model.

Figure A.4 is as Figure A.3, but for control models estimated from surrogate datasets with shuffled SFPDs. The decoding power of models estimated from original datasets was significantly larger than that of models estimated from surrogate

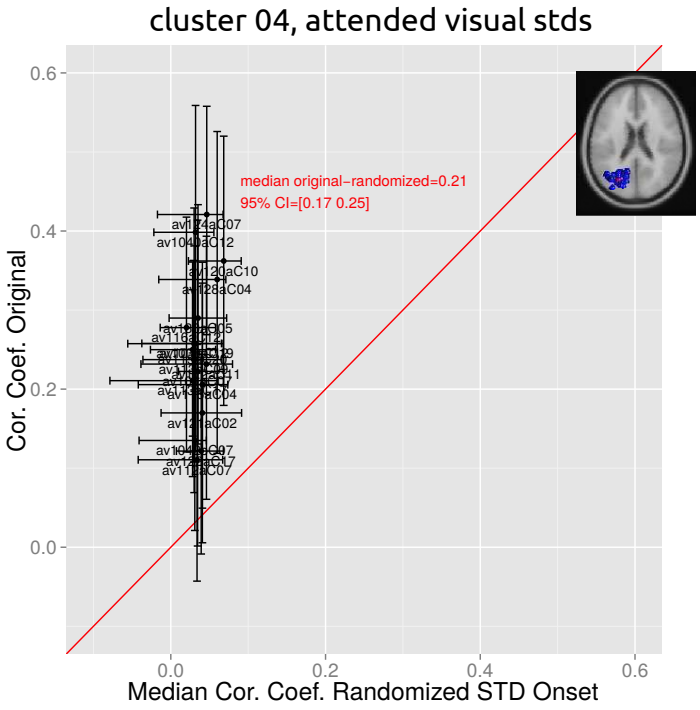


Figure A.3: Control for the influence of the warning signal on the SFP effect on ITPC. If the warning signal were generating the SFP effect on ITPC, randomizing the start times of the epochs should not significantly change the decoding power of models (see text). However, correlation coefficients between model decodings and experimental SFPDs for models fitted to original datasets, with epochs aligned to the presentation time of standards (ordinate), were significantly larger than those for models fitted to surrogate datasets, with epochs aligned to pseudo-random times (abscissa). Data is from the left parieto-occipital cluster 4 and attended visual standards. This control suggests that the warning signal is not generating the SFP effect on ITPC.

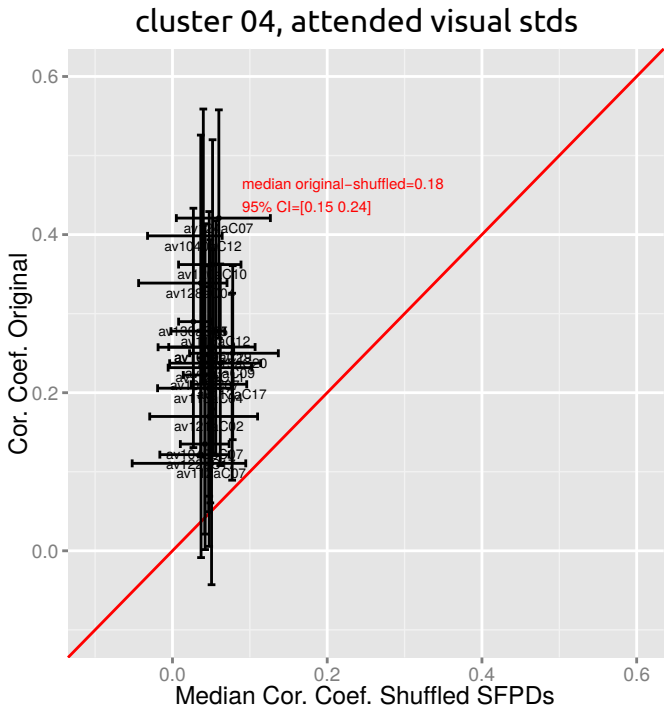


Figure A.4: Control suggesting that the SFPD generates the SFP effect on ITPC. Same format as in Figure A.3 but for surrogate datasets with SFPDs shuffled among trials. Correlations for models fitted to original datasets were significantly larger than those for models fitted to surrogate datasets, supporting the assumption that that the SFPD is the modulating variable in the SFP effect on ITPC.

datasets. The median pairwise difference between the correlation coefficient for a model estimated from an original dataset minus that for a model estimated from the corresponding surrogate dataset was 0.18, which was significantly different from zero (95% confidence interval: [0.15, 0.24]). Thus, Figure A.4 supports the inference that the SFPD is the experimental variable modulating the ITPC evoked by standards.

A.2.3 ERPs evoked by warning signals may occlude the SFP effect on ITPC

We hypothesized that the ERP evoked by the warning signal may complicate the observation of the SFP effect on ITPC and be a source of noise on this effect. To test this hypothesis we estimated models using only standards presented more than one second after the warning signal. Since the strength of the ERP varies across subjects and conditions (i.e., standard modalities and attended modalities), we thought that the hypothetical ERP noise could vary similarly. Thus, for each dataset (i.e., for each subject, standard modality and attended modality) we searched by brute force for an optimal minimum SFPD longer than one second and for an optimal maximum SFPD. We estimated decoding models for all combinations of minimum and maximum SFPDs (we used minimum and maximum SFPDs between one second and the maximum SFPD in each dataset in steps of 100 ms) and selected the optimal minimum and maximum SFPDs as the combination for which the correlation coefficient of the corresponding model was largest. For each combination of minimum and maximum SFPD Figure A.5a shows the correlation coefficient between model leave-one-out cross-validated decodings and SFPDs, for component 01 (assigned to the mid-central cluster 19) of subject av113a and unattended visual standards. Gray points in this plot correspond to combinations of minimum and maximum SFPDs for which there were not enough standards to estimate a model (Section 2.8) or for which the estimated mean correlation coefficient was negative (negative correlation coefficients can occur because the ridge-regression estimator is biased, Hoerl and Kennard, 1970). Best decodings were obtained for minimum and maximum SFPDs of 3,100 and 5,600 milliseconds, respectively, with a correlation coefficient $r = 0.83$.

Figure A.5b plots the estimated minimum versus maximum SFPDs, maximum correlation coefficients and 95% confidence intervals for all subjects and

components in the mid-central cluster 19 and unattended visual standards (the point highlighted in red corresponds to the subject and component shown in Figure A.5a). Ten of the eleven estimated models with minimum SFPD selected to be larger than one second decoded SFPDs significantly better than chance (Figure A.5b). The mean correlation coefficient for the five models with larger minimum SFPD was significantly larger than that of the six models with smaller minimum SFPD (difference in mean correlation = 0.33, p=0.002; permutation test for difference in means). The decoding power of these models was significantly larger than that of models estimated using the original procedure (i.e., with minimum SFPD set to zero) (median correlation coefficient difference = -0.30, 95% confidence interval = [-0.39, -0.06]; Figure A.5c), while the decoding power of models with minimum SFPD selected to be larger than one second remained significantly correlated with error rates of subjects (r=-0.61, p=0.0215; Figure A.5d).

That models with minimum SFPD selected to be larger than one second achieved larger decoding power than original models with minimum SFPD set to zero (Figure A.5c) could be due to the fact that we searched over a larger set of models to select the optimal minimum and maximum SFPD for models with minimum SFPD selected to be larger than one second. However, this difference in the size of the model search space cannot explain the larger decoding power of models with larger minimum SFPD (Figure A.5b), since the size of the model search space was the same for all estimated models. The larger decoding power of models with minimum SFPD selected to be larger than one second should not reflect overfitting, because decoding power was measured using leave-one-out crossvalidation and the estimation of model coefficients was heavily regularized. Also, if the larger decoding power of models with minimum SFPD selected to be larger than one second were an artifact, it would be unlikely to obtain significant correlations between the decoding power of models and error rates of subjects (as

in Figure A.5d).

We conjecture that superior decodings occurred because, by removing standards close to the warning signal, we eliminated uninformative (i.e., noisy) training samples for the estimation of decoding models. This late finding suggests that the results presented in this manuscript may be further improved by selecting minimum and maximum SFPDs for each subject and condition (as in Figure A.5a) instead of just selecting maximum SFPDs (as in Figure A.2).

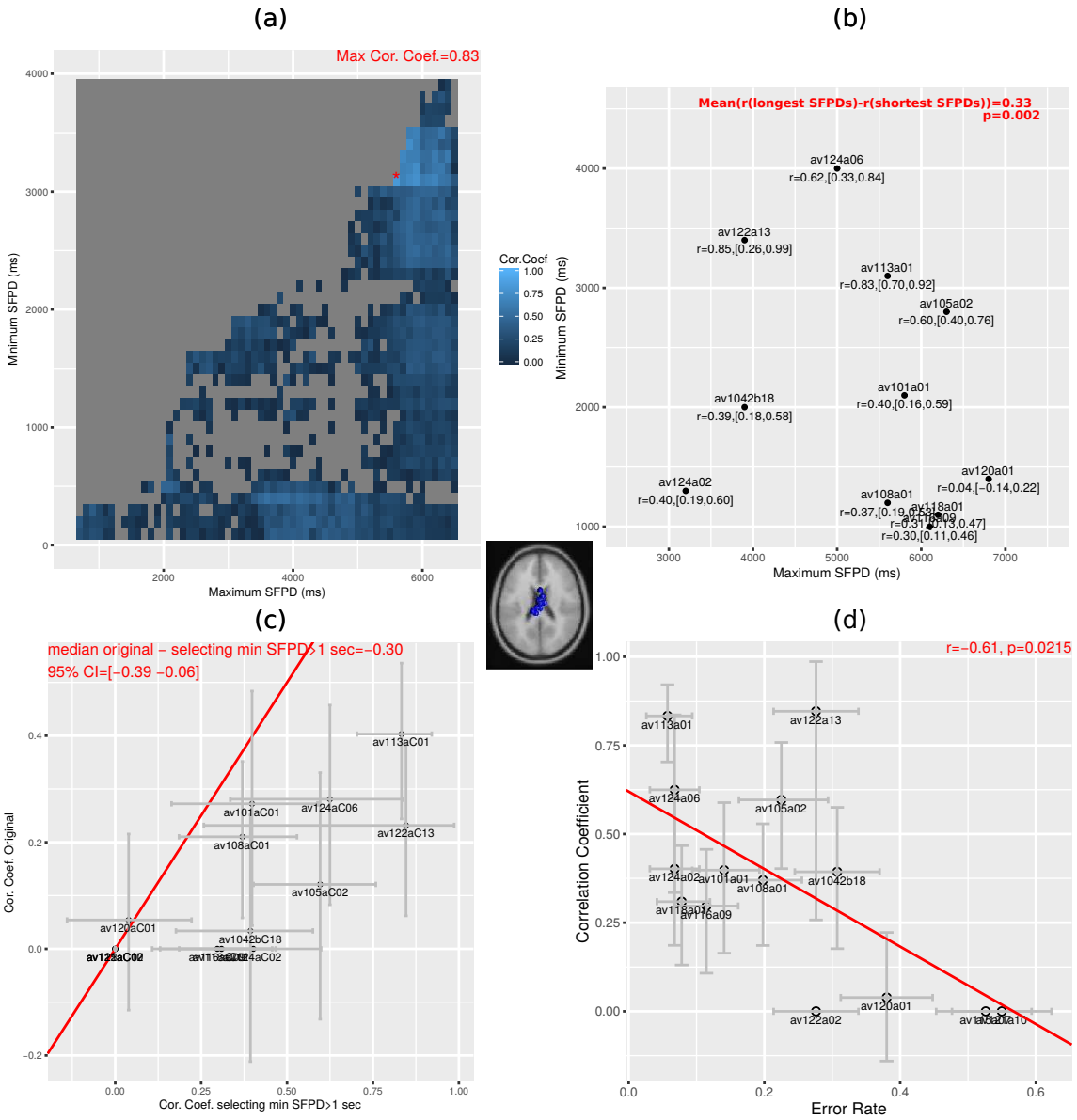


Figure A.5: (From previous page) Selection of minimum SFPDs larger than one second for subjects and components in the mid-central cluster 19 and unattended visual standards. (a) Example selection for subject av113a, component 01 and unattended visual standards. We estimated decoding models with minimum and maximum SFPDs varying from 1000 ms to the maximum possible SFPD of 6,612 ms in steps of 100 ms and selected optimal minimum and maximum SFPDs as those corresponding to the model with largest correlation coefficient. For this subject, component and condition the maximum correlation coefficient was r=0.83, which was achieved by a decoding model estimated with minimum and maximum SFPDs of 3,100 and 5,600 milliseconds, respectively. (b) Minimum versus maximum optimal SFPDs for all models, their corresponding correlation coefficients and 95% confidence intervals. The correlation coefficients of ten of the eleven estimated models were significantly different from zero. The mean correlation coefficient of the five models with the longest minimum SFPD was significantly larger than that of the six models with the shortest minimum SFPD (difference of means=0.33, P value=0.002). (c) Decodings of models estimated using the original procedure (i.e., using a minimum SFPD of zero and selecting the maximum SFPD, as in Figure A.2) versus those of models estimated using minimum SFPDs selected to be larger than one second. On average, the latter set of models decoded SFPDs significantly better than the former set of models (median correlation coefficient for models estimated from original datasets minus that for models estimated from datasets with minimum SFPD selected to be larger than one second = -0.30, 95% confidence interval = [-0.39, -0.06]). (d) Decoding power of models with minimum SFPD selected to be larger than one second was significantly correlated with subjects' error rates (r=-0.61, p<=0.0215).

A.3 How may the brain generate the SFP effect on ITPC?

A brain mechanism generating the SFP effect on ITPC should be able to replicate the fluctuations in ITPC shown in Figures 3 and A.11. We make three observations about these figures: First, these fluctuations appear to oscillate at a very low frequency (i.e., less than 1 Hz), independently of the frequency at which the phase of trials was measured (the latter frequency is that of the cosine of the mean phase given by the dashed curve in Figures 3d-f and A.11). Coincidentally, fluctuations at 1 Hz in visual detectability have been reported by Fiebelkorn et al. (2011). Also, fluctuations in somatosensory detectability (between 0.01 and 0.1 Hz) have been described by Monto et al. (2008). Second, in some figures the phase of the oscillations of ITPC at time zero differs between early and late trials (e.g., the phase of the blue curve corresponding to early trials in Figure 3d is more advanced than that of the red curve corresponding to late trials). Third, as shown in Figure 5, these fluctuations in ITPC are related to subjects' behaviors.

We propose a simple model for the generation of low-frequency oscillations that can account for the previous observations. The component of the recorded potential corresponding to trial i, at frequency f, and time t is given by Eq. A.12:

$$trial(i,f,t) = cos(2\pi ft + \theta + n(SFPD[i],t)), \text{ with} \quad (A.12)$$

$$n(SFPD,t) = \mathcal{M}(\mu = 0, \kappa = \kappa(SFPD,t)) \quad (A.13)$$

$$\kappa(SFPD,t) = \frac{\max\kappa - min\kappa}{2} cos(2\pi f_n t + \theta(SFPD)) + \frac{min\kappa + \max\kappa}{2}, \text{ and} \quad (A.14)$$

$$\theta(SFPD) = \frac{\pi}{(\text{maxSFPD} - \text{minSFPD})^s}(\text{maxSFPD} - SFPD)^s \quad (A.15)$$

To account for the first previous observation, we assume that the phase of the cosine is contaminated by an additive noise n (Eq. A.13) following a von Mises distribution with a precision parameter, κ, varying sinusoidally on time (Eq. A.14).

When the precision of this noise is small and large we should observe decoherent and coherent oscillations, respectively, and, because the noise precision varies sinusoidally in time (Equation A.14), we should observe alternations between coherence and decoherence, as shown in Figures 3d-f and A.11. To account for the second previous observation, we make the phase at time zero of the precision of the noise vary smoothly as a function of the SFPD (Eq. A.15). To account for the third observation, we speculate that subjects achieving better detection performance were those showing larger fluctuations of the precision of the noise. For these subjects ITPC should be more different between trials closer to and further away from the warning signal, and models should more reliably decode SFPDs from ITPC, in agreement with Figure 5. The parameters of the previous model are the frequency, f, the noiseless phase, θ, and the vector of SFPDs, $SFPD$, in Eq. A.12; the minimum and maximum values, $min\kappa$ and maxκ, and the frequency, f_n, of the noise precision in Eq. A.14; and the steepness of the change in the noise precision phase as a function of the SFPD, s, in Eq. A.15.

To validate the previous speculations we simulated data following the model in Eq. A.12 (Section A.4, Figure A.7) and fitted a decoding model to this data, as we did with the neural recordings (Section 2.7). With this simulated data we obtained oscillations of mean DMP (Figures A.6c,d, as in Figure 3d-f), in the difference between the averaged DMP between trials closer to and further away from the warning signal (Figures A.6e,f, as in Figure 3g-i), and in estimated model coefficients (Figures A.6g,h, as in Figure 3m-o). In addition, we verified that a model fitted to data with larger fluctuations of the precision of the noise generated more accurate decodings than those of a model fitted to data with smaller modulations of this precision, supporting the previous argument on the relation between fluctuation of ITPC and subjects' behaviors.

These results show that a simple sinusoidal oscillation with a noise process added to its phase captures salient features of the SFP effect on ITPC, suggesting

that this model may be a good first approximation to how the SFPD modulates the ITPC of neural oscillations. Key features of this noise are its precision oscillating in time with a phase depending on the SFPD.

A.4 Simulations of the SFP effect on ITPC

We simulated the model in Eq. A.12 twice, with the precision of the phase noise n displaying larger (maxκ=5.00, minκ=0.01) and smaller (maxκ=3.50, minκ=1.51) oscillations and with parameters $f = 7\,Hz$, $\theta = \pi$, $f_n = 3\,Hz$, and $s = 2$. For these simulations we used the SFPDs, $SFPD$ in Eq. A.12, from one experimental session (subject av124a and attended visual standards). The simulated oscillations are shown in Figure A.7 as an erpimage (Makeig et al., 2004) sorted by SFPDs (black curve to the left of time zero). Figure A.7a and A.7b correspond to larger and smaller fluctuations in the precision of the phase noise, respectively. Figure A.6 is as Figure 3 but for the simulated data. The fluctuations in averaged DMP in trials closest to and furthest from the warning signal, those in their difference, and fluctuations in the estimated model coefficients are similar for simulated and experimental data. The correlation coefficient for the model fitted to data with smaller fluctuations of the precision of the noise was $r = 0.23$ (95% CI=[0.12, 0.34]), which was significantly smaller than that for the model fitted to data with larger fluctuations of the precision of the noise, $r = 0.51$ (95% CI=[0.43, 0.59]).

A.5 Problematic 2D visualization of 3D ECDs

Our study only characterized ICs with equivalent current dipole (ECD) inside the brain (Section A.1.1 ICA data preprocessing). The appearance that some ECDs lie outside the brain is due to the difficult two-dimensional visualization of three-dimensional ECDs.

Some ECDs appear to lie outside the brain because it is difficult to display

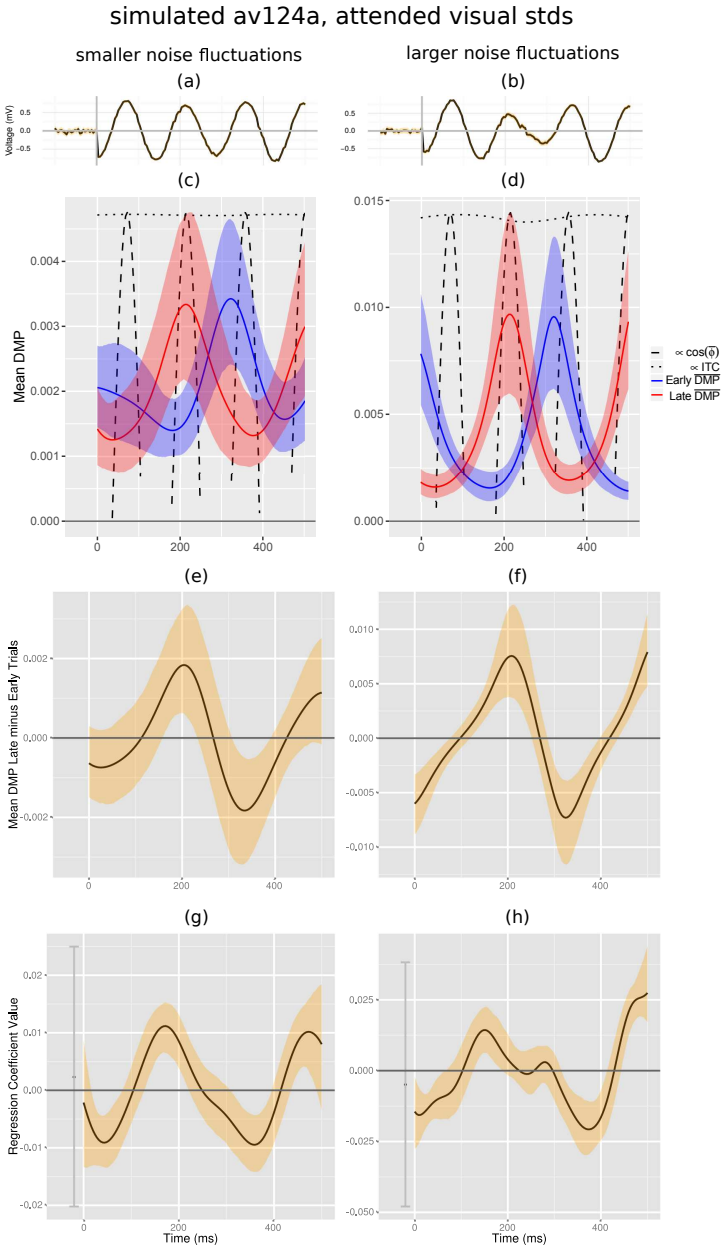


Figure A.6: Average DMP in trials furthest from and closest to the warning signal and regression coefficients for data simulated from Eq. A.12. Same format as Figure 3. That the panels in this figure capture salient features of those in Figure 3, generated from neural recordings, suggests that the simple oscillatory model in Eq. A.12 is a good first approximation to how the SFPD modulates the phase of neural oscillations.

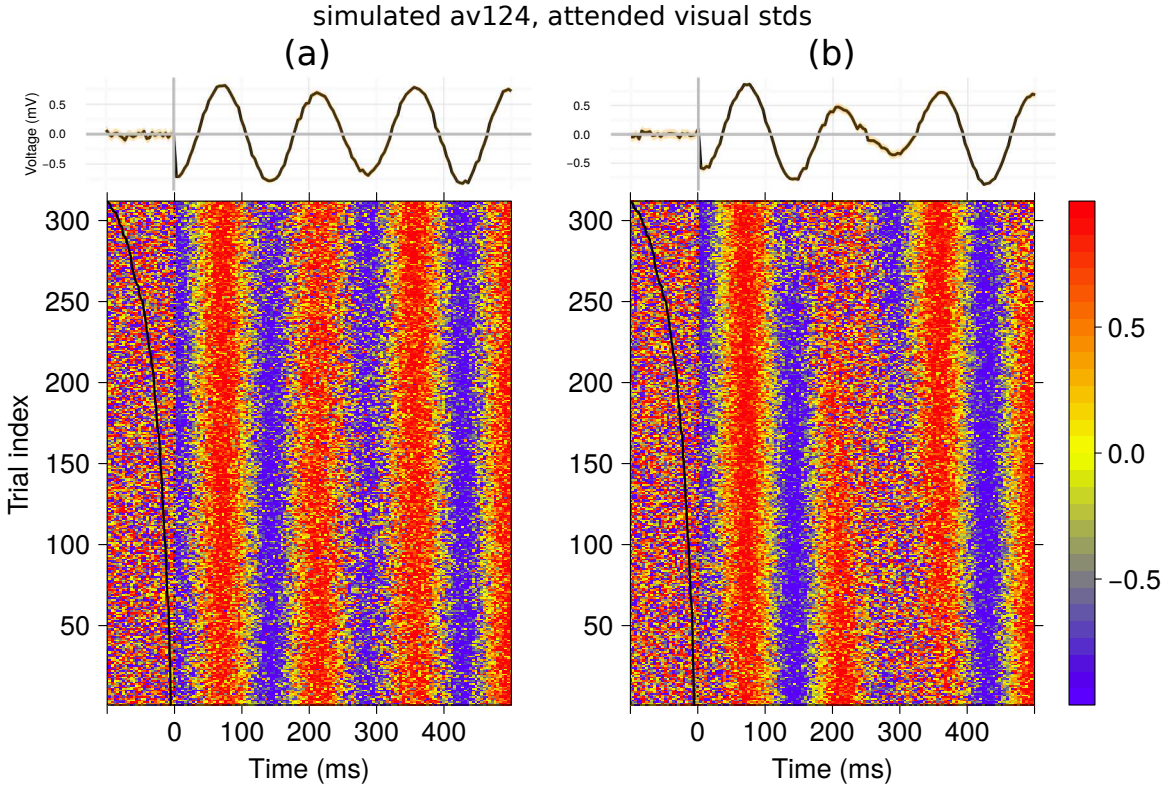


Figure A.7: Oscillations simulated according to Eq. A.12 with phase noise n having small ($min\kappa = 1.51$, $max\kappa = 3.5$, panel (a)) and large ($min\kappa = 0.01$, $max\kappa = 5.0$, panel (b)) fluctuations in precision (Eq. A.14). Bottom panels: each row represents a trial color coded according the simulated voltage; trials are sorted from bottom to top by increasing SFPD. We used the SFPDs from subject av124a and attended visual standards. The black curves at negative times plot the presentation times of the warning signal scaled so that the earliest presentation time fits in the 100 ms-long pre-stimulus window. Top panels: ERPs calculated from trials in the bottom panels.

three-dimensional ECDs locations on two-dimensional images. Figure 4 shows projections of ECDs locations on an axial slice of the brain at $z = 0$, while Figure A.9 displays projections of ECDs on axial, coronal, and sagital slices at $z = 0$, $x = 0$, and $y = 0$, respectively. It is not possible to judge from these projections if an ECD is located in gray matter, in white matter, or in the ventricles. For example, an ECD at location (x, y, z) could be in gray matter but the projected location $(x, y, 0)$ could be in the ventricles, thus in its axial projection (as in Figure 4 and the left panels in Figure A.9) the ECD would appear to be located in the ventricles. In addition, the previous figures display brain slices from the Montreal Neurological Institute (MNI) standard brain that does not exactly represent the brain of any subject in our study. Therefore, one should only expect that the location of an ECD should be close to, but not exactly in, gray matter.

A.6 Supplementary figures

Figure A.8 plots examples of deviant foreperiod effects on reaction times and detectability. Figure A.9 plots axial, coronal, and sagital views of all clusters. Figure A.10 shows the number of significant models estimated for each subject, standard modality and attended modality as a function of subjects' error rates. Figure A.11 shows average DMPs for the 20% trials with the shortest and longest SFPDs in ICs from the left parieto-occipital cluster 4. Figure A.12 shows that the ERP variability at its peak value and the peak ITC value are both uncorrelated with subjects' error rates for the mid-central cluster 19 and unattended visual standards.

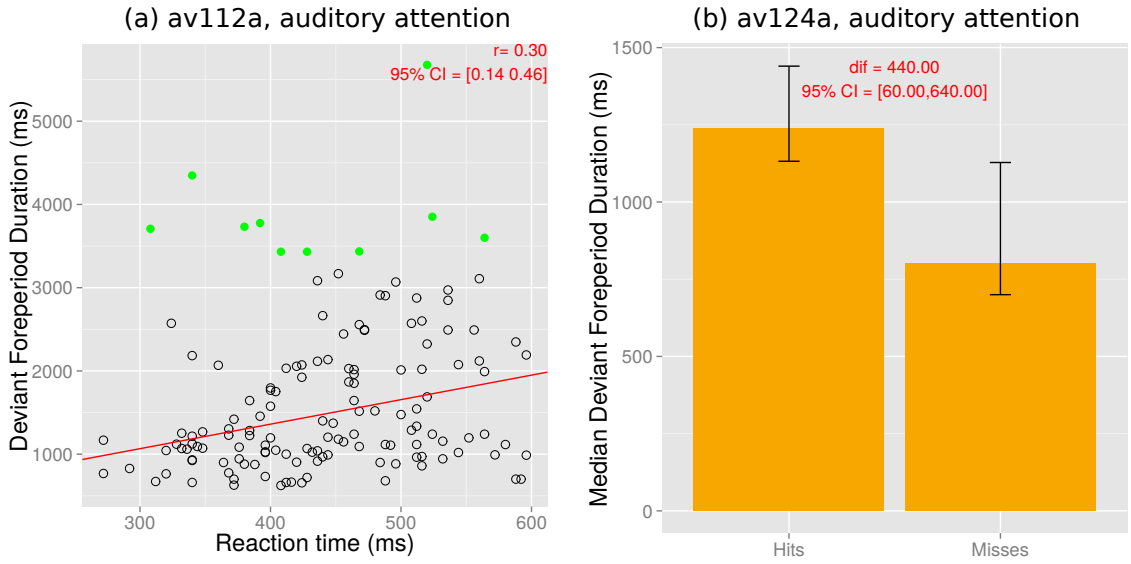


Figure A.8: Examples of significant deviant foreperiod effects on reaction time (a) and on stimulus detectability (b). (a) Deviant foreperiod durations as a function of reaction times, for subject av112a and the auditory attended modality. The significant correlation coefficient indicates a deviant foreperiod effect on reaction times. Points colored in green mark outliers detected in the computation of the robust correlation coefficient (Section A.1.10). (b) Median deviant foreperiod duration for hits and misses, for subject av124a and the auditory attended modality. The significantly larger median foreperiod duration for hits than for misses indicates a foreperiod effect on stimulus detectability.

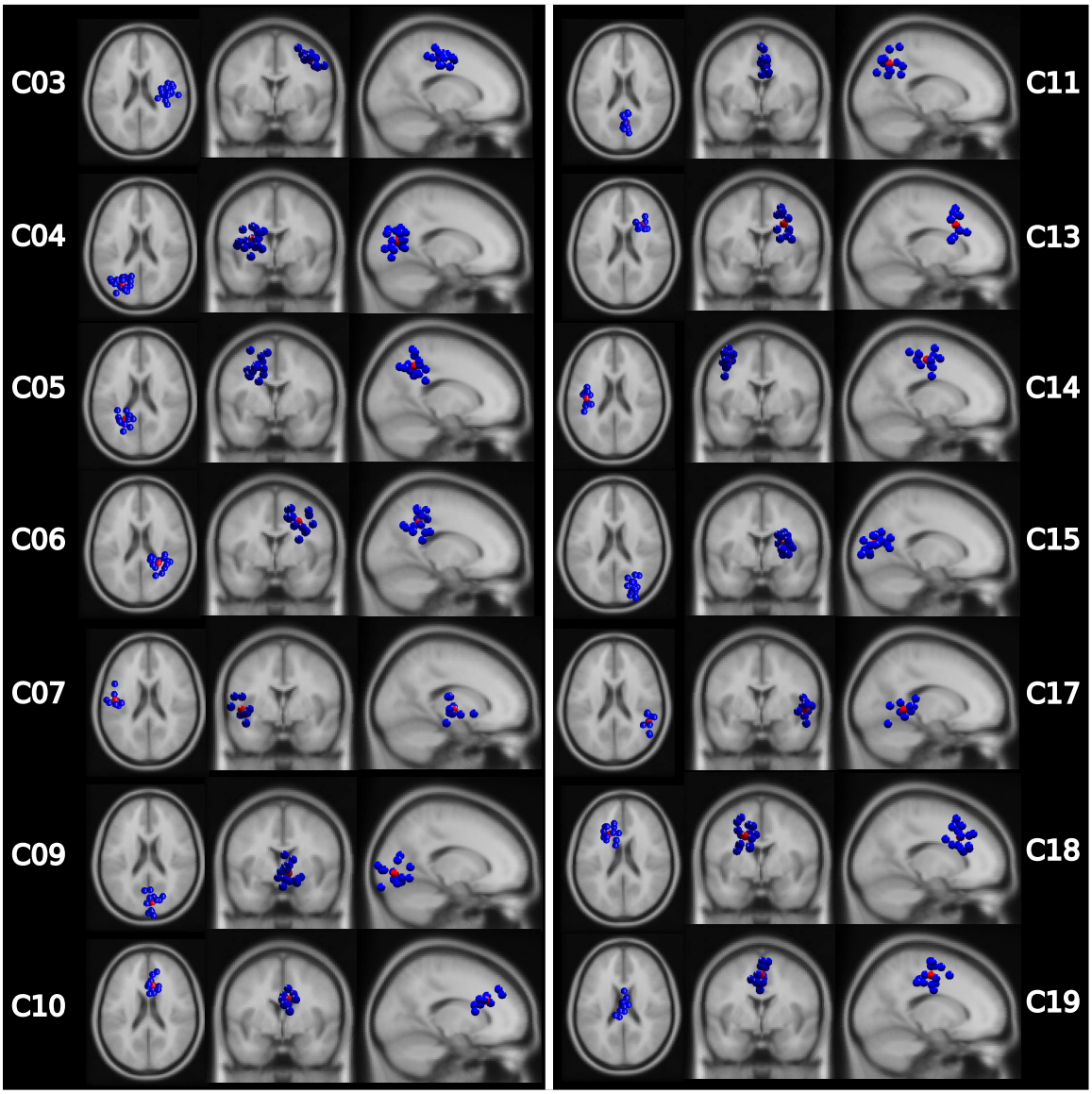


Figure A.9: Clusters of ICs. Left, center, and right columns correspond to axial, coronal, and sagital views, respectively.

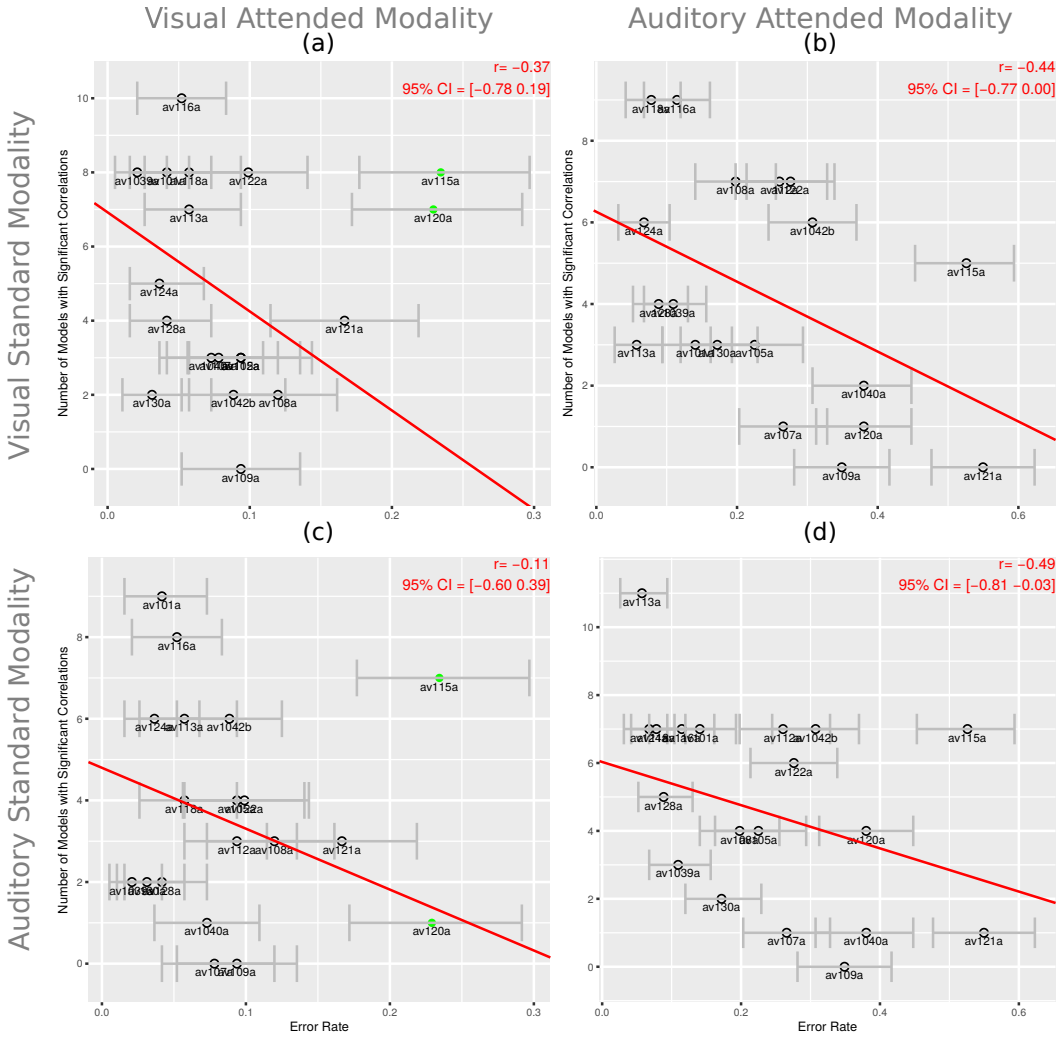
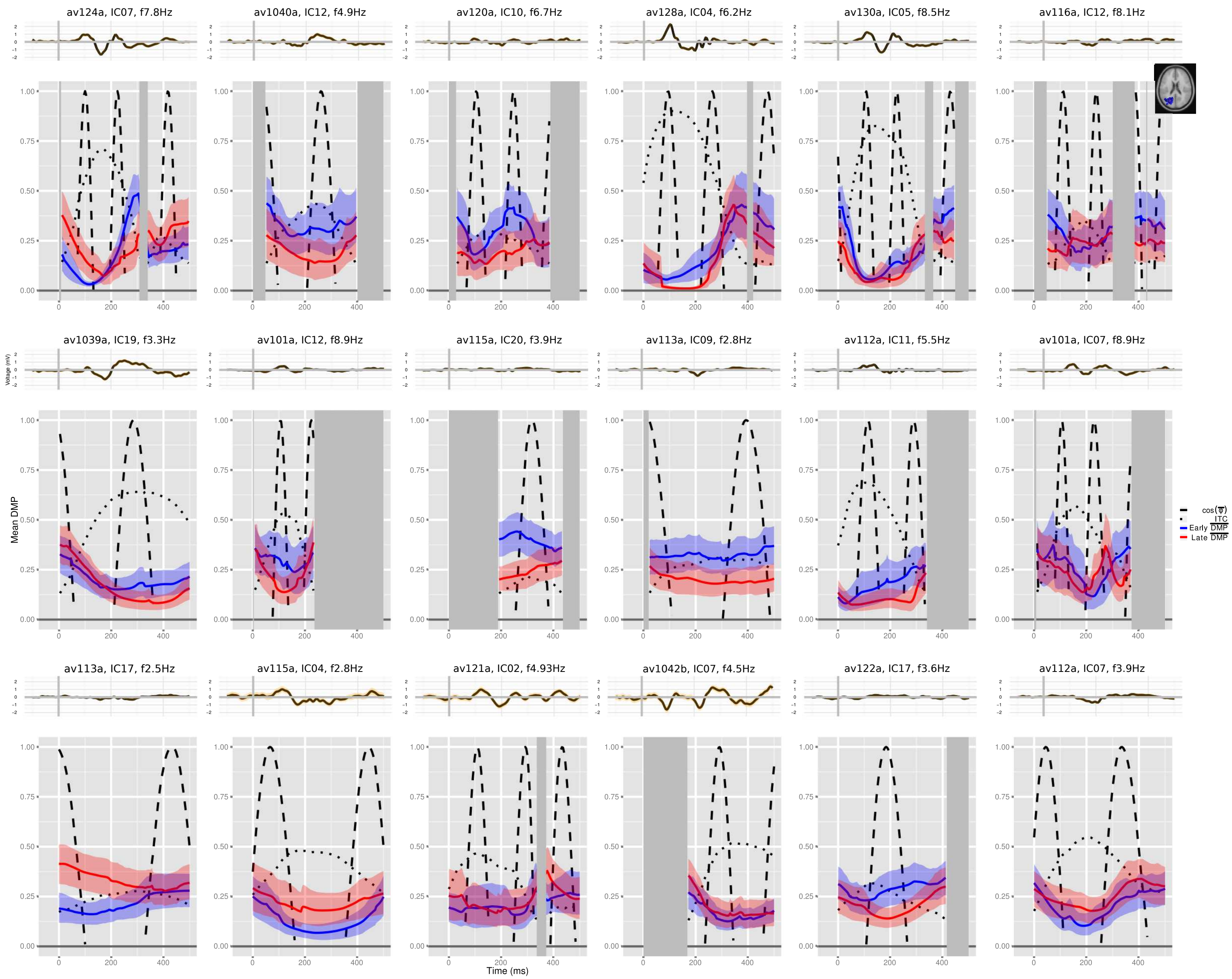


Figure A.10: Distribution of the SFP on ITPC among subjects. For each subject panels plot the number of models with significant correlations between models' decoding and experimental SFPDs (as in Figure 3g-i) versus the subject error rate, for the visual and auditory standard modalities (top and bottom rows, respectively) and for the visual and auditory attended modalities (left and right columns, respectively). The SFP on ITPC was evident in most subjects and for the auditory attended modality it tended to be stronger for subjects attaining lower error rates. Green points indicate outliers detected in the calculation of robust correlation coefficients (Section A.1.10).


av124a, IC07, f7.8Hz
av1040a, IC12, f4.9Hz
av120a, IC10, f6.7Hz
av128a, IC04, f6.2Hz
av130a, IC05, f8.5Hz
av116a, IC12, f8.1Hz
av1039a, IC19, f3.3Hz
av101a, IC12, f8.9Hz
av115a, IC20, f3.9Hz
av113a, IC09, f2.8Hz
av112a, IC11, f5.5Hz
av101a, IC07, f8.9Hz
av113a, IC17, f2.5Hz
av115a, IC04, f2.8Hz
av121a, IC02, f4.93Hz
av1042b, IC07, f4.5Hz
av122a, IC17, f3.6Hz
av112a, IC07, f3.9Hz
Voltage (mV)
Mean DMP
Time (ms)
ITC
Early DMP
Late DMP

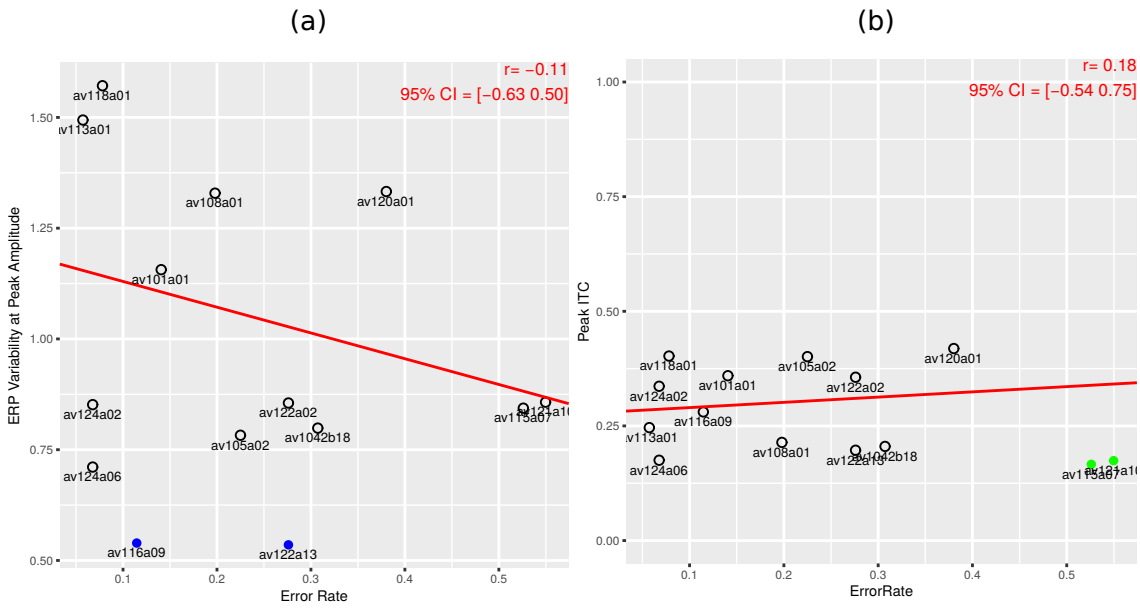


Figure A.12: Additional evidence for the lack of correlation between features of the warning signal ERP and subjects' error rates in the mid-central cluster 19 and unattended visual standards. (a) Variability of the ERP at its peak value versus subjects' error rates. Variability was quantified as the width of the 95% confidence interval of the mean of trials. (b) Peak ITC value versus subjects' error rates.

Figure A.11: (From previous page) Mean DMP for trials with the shortest and longest SFPDs. Each panel corresponds to a different IC from the left parieto occipital cluster 4 and attended visual standards; the top plot shows the ERP from all trials and the bottom the average DMPs from the 20% trials with the shortest (blue traces) and longest (red traces) SFPDs. Panels are sorted from left to right and from top to bottom by decreasing decoding power of the corresponding models. Dotted and dashed lines plot the ITC and the cosine of the mean phase (Equation A.6), respectively. For most ICs there is a significant SFP effect on ITPC (i.e., a significant difference between the red and blue curves).

Table A.1: Information about clusters of ICs. The anatomical labels associated with the Talairach coordinates of the clusters' centroids were extracted using the Talairach client (Lancaster et al., 1997, 2000)

Cluster	No. Subjects	No. ICs	Talairach X	Y	Z	Hemisphere	Lobe	Gyrus	Brodmann Area
3	12	17	45	-13	43	Right	Frontal Lobe	Precentral Gyrus	4
4	14	18	-19	-74	9	Left	Occipital Lobe	Cuneus	17
5	11	14	-14	-42	33	Left	Limbic Lobe	Cingulate Gyrus	31
6	12	17	43	-39	34	Right	Parietal Lobe	Supramarginal Gyrus	40
7	8	9	-38	-18	1	Left	Sub-lobar	Claustrum	*
8	3	4	56	-1	12	Right	Frontal Lobe	Precentral Gyrus	6
9	12	14	3	-62	7	Right	Limbic Lobe	Posterior Cingulate	30
10	9	10	6	17	18	Right	Limbic Lobe	Anterior Cingulate	33
11	12	13	8	-46	28	Right	Parietal Lobe	Precuneus	31
12	3	4	7	18	-50	No Gray Matter found			
13	11	12	30	10	34	Right	Frontal Lobe	Middle Frontal Gyrus	8
14	10	13	-35	-25	53	Left	Frontal Lobe	Precentral Gyrus	4
15	16	20	23	-87	10	Right	Occipital Lobe	Middle Occipital Gyrus	18
16	4	5	14	45	-3	Right	Limbic Lobe	Anterior Cingulate	32
17	7	9	48	-33	4	Right	Temporal Lobe	Superior Temporal Gyrus	22
18	12	15	-6	21	31	Left	Limbic Lobe	Cingulate Gyrus	32
19	12	14	2	-7	42	Right	Limbic Lobe	Cingulate Gyrus	24

A.7 Supplementary tables

Table A.1 provides the Talairach coordinates and anatomical labels of the centroids of the clusters in Figure 4. For each cluster, standard modality, and attended modality, Table A.2 gives the number of models with decodings significant correlated with SFPDs, and Tables A.3 and A.4 give the correlation coefficient, r, between the decoding power of models and subjects' detection error rates and mean reaction times, respectively, as well as the corresponding unadjusted, p, and adjusted, adj_p, P values.

Table A.2: Number (n) and proportion (%) of models with significant correlations (adj_p<0.05) between models' decodings and experimental SFPDs. Cells highlighted in blue correspond to clusters, standard modalities, and attended modalities where for more than 40% of models decodings were significantly correlated with SFPDs.

Cluster	Standard Modality	Visual Attention		Auditory Attention	
		n	%	n	%
03	Visual	05	0.29%	06	0.35%
	Auditory	02	0.12%	02	0.12%
04	Visual	08	0.44%	08	0.44%
	Auditory	02	0.11%	05	0.28%
05	Visual	04	0.29%	02	0.14%
	Auditory	01	0.07%	03	0.21%
06	Visual	04	0.24%	03	0.18%
	Auditory	05	0.29%	06	0.35%
07	Visual	01	0.11%	01	0.11%
	Auditory	00	0.00%	03	0.33%
09	Visual	06	0.43%	05	0.36%
	Auditory	03	0.21%	05	0.36%
10	Visual	01	0.10%	02	0.20%
	Auditory	00	0.00%	05	0.50%
11	Visual	06	0.46%	03	0.23%
	Auditory	03	0.23%	04	0.31%
13	Visual	03	0.25%	02	0.17%
	Auditory	03	0.25%	03	0.25%
14	Visual	04	0.31%	05	0.38%
	Auditory	04	0.31%	02	0.15%
15	Visual	11	0.55%	07	0.35%
	Auditory	01	0.05%	07	0.35%
17	Visual	02	0.22%	02	0.22%
	Auditory	03	0.33%	03	0.33%
18	Visual	01	0.07%	03	0.20%
	Auditory	06	0.40%	03	0.20%
19	Visual	04	0.29%	02	0.14%
	Auditory	03	0.21%	04	0.29%

Table A.3: Correlations between the strength of the SFP effect on ITPC and subjects' detection error rates. The strength of the SFP effect on ITPC is quantified with the correlation coefficient between decodings of the model and experimental SFPDs (Section 3.3). Each cell shows the correlation coefficient, r, and P values unadjusted, p, and adjusted, adj_p, for multiple comparisons. Blue cells highlight correlations with p<0.05 and the red cell corresponds to adj_p=0.05.

Cluster	Standard Modality	Visual Attention r	Visual Attention p	Visual Attention adj_p	Auditory Attention r	Auditory Attention p	Auditory Attention adj_p
03	Visual	0.29	0.3054	1.00	0.30	0.2714	1.00
	Auditory	-0.07	0.8022	1.00	-0.20	0.4766	1.00
04	Visual	-0.64	0.0048	0.38	0.16	0.5212	1.00
	Auditory	0.06	0.8360	1.00	-0.12	0.6572	1.00
05	Visual	-0.50	0.0916	1.00	-0.70	0.0162	0.77
	Auditory	-0.22	0.4846	1.00	-0.10	0.7612	1.00
06	Visual	-0.54	0.0336	0.95	-0.33	0.2068	1.00
	Auditory	0.15	0.5650	1.00	-0.23	0.3694	1.00
07	Visual	-0.74	0.0444	0.98	0.36	0.3490	1.00
	Auditory	-0.17	0.6548	1.00	-0.59	0.1016	1.00
09	Visual	-0.20	0.4902	1.00	-0.43	0.1456	1.00
	Auditory	-0.34	0.2412	1.00	-0.25	0.3942	1.00
10	Visual	-0.29	0.4274	1.00	0.49	0.1706	1.00
	Auditory	-0.18	0.6250	1.00	-0.05	0.8794	1.00
11	Visual	0.47	0.1218	1.00	-0.29	0.3104	1.00
	Auditory	-0.43	0.1388	1.00	-0.34	0.2586	1.00
13	Visual	-0.19	0.5626	1.00	0.26	0.4390	1.00
	Auditory	-0.23	0.5264	1.00	0.11	0.7402	1.00
14	Visual	-0.20	0.4964	1.00	0.34	0.2938	1.00
	Auditory	0.36	0.2404	1.00	-0.21	0.5538	1.00
15	Visual	0.23	0.3648	1.00	-0.08	0.7376	1.00
	Auditory	-0.38	0.1298	1.00	-0.09	0.7108	1.00
17	Visual	0.06	0.8684	1.00	-0.29	0.4408	1.00
	Auditory	0.11	0.7900	1.00	-0.08	0.8194	1.00
18	Visual	-0.10	0.7612	1.00	-0.11	0.7054	1.00
	Auditory	0.07	0.8038	1.00	-0.25	0.3922	1.00
19	Visual	-0.07	0.8098	1.00	-0.91	0.0004	0.05
	Auditory	-0.62	0.0216	0.85	-0.07	0.8002	1.00

Table A.4: Correlations between the strength of the SFP effect on ITPC and subjects' mean reaction times. Same format as Table A.3.

Cluster	Standard Modality	Visual Attention r	Visual Attention p	Visual Attention adj_p	Auditory Attention r	Auditory Attention p	Auditory Attention adj_p
03	Visual	-0.30	0.2574	1.00	0.02	0.9392	1.00
	Auditory	-0.25	0.3218	1.00	0.16	0.5410	1.00
04	Visual	-0.15	0.6102	1.00	0.22	0.3654	1.00
	Auditory	-0.04	0.8712	1.00	0.14	0.5598	1.00
05	Visual	0.13	0.6412	1.00	-0.57	0.0318	0.84
	Auditory	0.09	0.7596	1.00	-0.24	0.4150	1.00
06	Visual	-0.15	0.5712	1.00	-0.14	0.5892	1.00
	Auditory	0.48	0.0534	0.96	-0.16	0.5414	1.00
07	Visual	0.20	0.5940	1.00	-0.12	0.7456	1.00
	Auditory	0.11	0.7720	1.00	-0.03	0.9420	1.00
09	Visual	-0.01	0.9686	1.00	0.09	0.7458	1.00
	Auditory	0.13	0.6486	1.00	0.11	0.6886	1.00
10	Visual	-0.40	0.2706	1.00	0.54	0.1098	1.00
	Auditory	0.36	0.3294	1.00	-0.12	0.7308	1.00
11	Visual	0.13	0.6674	1.00	-0.02	0.9430	1.00
	Auditory	-0.45	0.1306	1.00	0.14	0.6354	1.00
13	Visual	0.12	0.6910	1.00	0.58	0.0524	0.95
	Auditory	-0.01	0.9682	1.00	0.35	0.2622	1.00
14	Visual	-0.34	0.2598	1.00	-0.03	0.9210	1.00
	Auditory	-0.25	0.4040	1.00	-0.07	0.7996	1.00
15	Visual	-0.04	0.8536	1.00	-0.00	0.9890	1.00
	Auditory	-0.07	0.7526	1.00	-0.06	0.8010	1.00
17	Visual	-0.23	0.5416	1.00	-0.07	0.8538	1.00
	Auditory	0.03	0.9172	1.00	-0.10	0.7754	1.00
18	Visual	-0.56	0.0270	0.80	0.14	0.6256	1.00
	Auditory	0.01	0.9728	1.00	-0.13	0.6502	1.00
19	Visual	-0.57	0.0336	0.86	-0.35	0.2044	1.00
	Auditory	-0.16	0.5780	1.00	0.18	0.5258	1.00