Noninvasively decoding the contents of visual working memory in the human prefrontal cortex within high-gamma oscillatory patterns

Polania, Rafael ; Paulus, Walter ; Nitsche, Michael A

DOI: https://doi.org/10.1162/jocn_a_00151

Noninvasively Decoding the Contents of Visual Working Memory in the Human Prefrontal Cortex within High-gamma Oscillatory Patterns

Rafael Polanía, Walter Paulus, and Michael A. Nitsche

Abstract

■ The temporal maintenance and subsequent retrieval of information that no longer exists in the environment is called working memory. It is believed that this type of memory is controlled by the persistent activity of neuronal populations, including the prefrontal, temporal, and parietal cortex. For a long time, it has been controversially discussed whether, in working memory, the PFC stores past sensory events or, instead, its activation is an extramnemonic source of top–down control over posterior regions. Recent animal studies suggest that specific information about the contents of working memory can be decoded from population activity in prefrontal areas. However, it has not been shown whether the contents of working memory during the delay periods can be decoded from EEG recordings in the human brain. We show that by analyzing the nonlinear dynamics of EEG oscillatory patterns it is possible to noninvasively decode with high accuracy, during encoding and maintenance periods, the contents of visual working memory information within high-gamma oscillations in the human PFC. These results are thus in favor of an active storage function of the human PFC in working memory; this, without ruling out the role of PFC in top–down processes. The ability to noninvasively decode the contents of working memory is promising in applications such as brain computer interfaces, together with computation of value function during planning and decision making processes. ■

INTRODUCTION

The use of information that is not present in the environment but is retained temporarily in memory systems until it is needed to guide behavior is called working memory. It is believed that this type of memory is controlled by the persistent activity of neuronal populations including the PFC. The crucial role of PFC during maintenance periods is unquestionable; however, it is controversially discussed whether PFC stores past sensory events or, instead, its activation is an extramnemonic source of top–down control over posterior regions (e.g., posterior parietal cortex [PPC] and anterior temporal cortex) that actually store the representations (Curtis & Lee, 2010; D'Esposito, 2007; Curtis & D'Esposito, 2003). Several fMRI studies have shown to be in favor of the role of PFC cortex along with PPC and anterior temporal cortex to participate in the active maintenance of working memory (Curtis, Rao, & D'Esposito, 2004; Cohen et al., 1997). However, other studies suggest that more posterior regions such as the PPC are centrally involved in active storage and manipulation processes, while PFC is related to the monitoring of the information that is being stored and manipulated by posterior regions (Champod & Petrides, 2007, 2010; Petrides, 2000).

Georg-August University of Göttingen

Using invasive recordings in animals, it was possible to decode from ensemble activities in PFC previous and future goal choices (Baeg et al., 2003). More recently, Averbeck and Lee (2007) combined invasive recordings and decoding methods to prove that the PFC neural activity of monkeys can retain information about sequences between trials. In a recent study combining MEG and multivariate decoding methods (Fuentemilla, Penny, Cashdollar, Bunzeck, & Duzel, 2010), the investigators provided evidence that maintenance in working memory is associated with replay of >13 Hz activity patterns of sensory input. Additionally, the authors show that such reactivations during maintenance periods are periodically modulated by theta oscillations at prefrontal and parieto-occipital regions. This study is in principle accordance with computational models, which suggest that one of the mechanisms enabling working memory is periodic reactivation of the maintained information coordinated by theta and gamma oscillations (Jensen & Lisman, 2005; Lisman, 1999). Regarding the mechanisms of the maintenance of information in the delay periods, one possibility is that the sustained neuronal activity during the delay periods is based on synaptic reverberation in recurrent circuits (Brunel & Wang, 2001; Wang, 2001). Other studies propose that each neuron does not have to maintain continuous neural activity throughout the delay periods, but working memory is mediated by sequential activation of different neuronal populations (Baeg et al., 2003), which are possibly

replayed during the whole delay period (Fuentemilla et al., 2010). If the latter is the case, it is plausible that this type of activity is reflected in scalp EEG recordings as oscillatory patterns, which might be captured by studying its nonlinear dynamics (Stam, Pijn, Suffczynski, & Lopes da Silva, 1999; see Methods section, for more details).

Thus, in the present study, we tested the hypothesis that PFC is involved in the active storage of information in the human brain, as previously suggested but not directly evidenced in the human, and moreover, we tested whether it is possible to noninvasively decode its contents by studying the nonlinear dynamics of high-frequency oscillatory patterns.

METHODS

Participants

Twelve healthy volunteers (eight women, aged 22–30 years) with normal or corrected-to-normal vision were included in the study. Subjects were informed about all aspects of the experiment, and all gave written informed consent. None of the participants suffered from any neurological or psychological disorder or took medication regularly or during the time the experiment was conducted. All subjects were right-handed, according to the Edinburgh handedness inventory (Oldfield, 1971). The experiments conform to the Declaration of Helsinki, and the experimental protocol was approved by the Ethics Committee of the University of Göttingen.

Stimuli

Subjects were seated in a comfortable armchair and kept their left and right index finger over push buttons. The computer monitor was placed in front of them at a distance of 80 cm. Participants performed a delayed letter discrimination task. Observers attended to two sample letters (''L'' and ''T,'' subtending 2° of visual angle) that were briefly presented (350 msec) in a randomized order. Immediately after each of the letters was presented, a mask stimulus displaying all possible line segments forming the letter stimulus L or T was presented for 1 sec to interrupt visual processing of the target shape. After the second mask was presented, a numerical cue (2°) indicated whether to remember the first or the second letter (Figure 1). After a

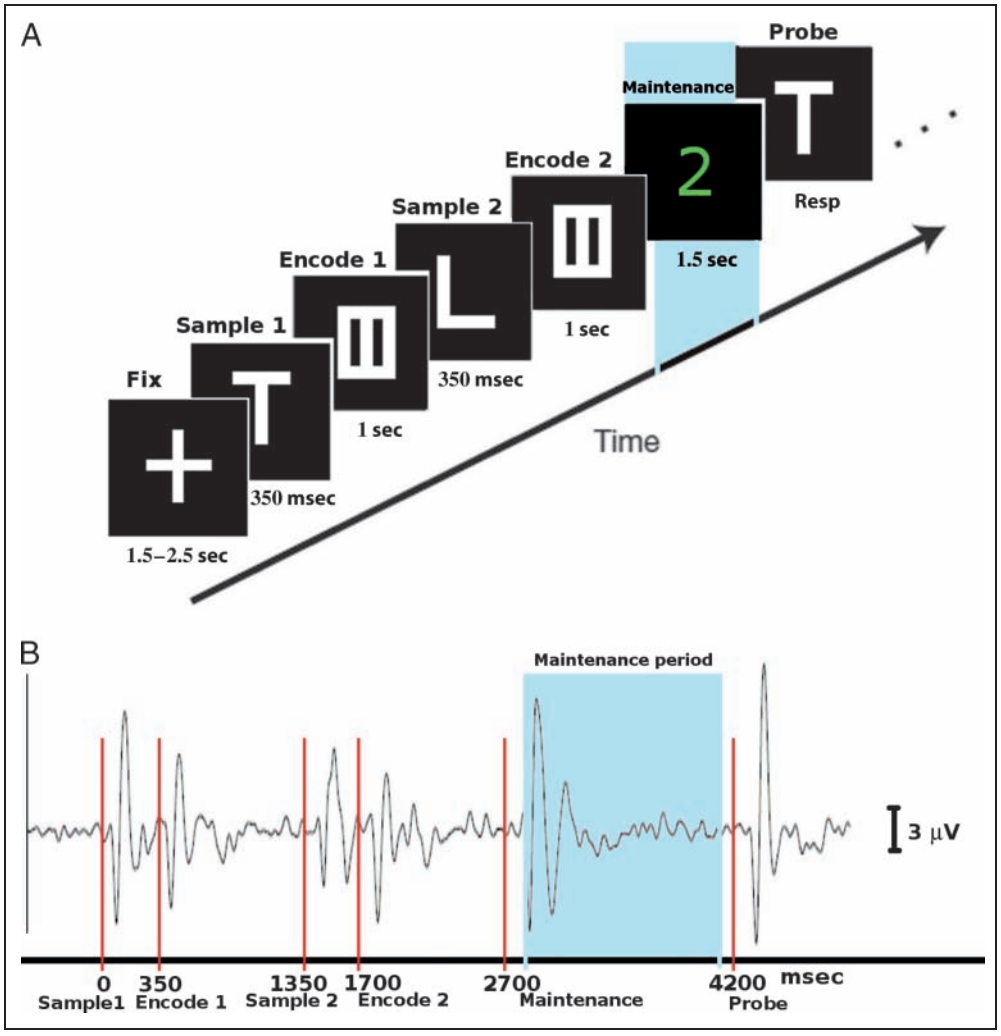

Figure 1. Design of the working memory experiment. (A) Timing events for an example working memory trial. Two letters (L and T) were briefly presented in randomized order, followed by a numerical cue (1 = *green* or 2 = *red*), indicating which letter to remember. After a 1.5-sec retention period, a test letter was presented, and subjects had to indicate whether it was L or T, relative to the cued letter. (B) For visualization purposes, we show the mean EEG activity in the visual cortex of one subject during the experiment, band-pass filtered at 1–40 Hz. Red lines show the starting point of each stimulus. Note the visual evoked potentials after each stimulus onset. Blue-shaded area shows the interval where activity patterns were intended to be decoded, that is, the maintenance period.

1.5-sec delay interval, a test letter was presented, and participants indicated as fast as possible whether it matched the numerically cued letter or not (right hand if it matched, left hand otherwise). By presenting the same two letters in every trial, we ensured that stimulus-driven activity could not predict the letter held in working memory. The memory cue appeared after the presentation of the stimuli to hold in memory and not beforehand, because otherwise subjects might attend more to the appearance of the cued letter (Harrison & Tong, 2009; Kamitani & Tong, 2005). Each experimental run contained four blocks, where each one consisted of 60 trials. Break between blocks was around 1 min.

EEG Recordings

EEGs were recorded against an average reference electrode using sintered Ag–AgCl electrodes (EEG-ANT system, the Netherlands) at the following 64 positions of the international 10–20 system: Fp1, Fpz, Fp2, F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, T7, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, Poz, O1, Oz, O2, AF7, AF3, AF4, AF8, F5, F1, F2, F6, FC3, Fcz, FC4, C5, C1, C2, C6, CP3, Cpz, CP4, P5, P1, P2, P6, PO5, PO3, PO4, PO6, FT7, FT8, TP7, TP8, PO7, PO8, M1, M2. Four additional channels were used: two horizontal EOG channels lateral to the left and right eyes' canthi and two vertical EOG channels, one below (suborbital) and one above (supraorbital) the right eye. Electrode impedance was monitored throughout the experiment to be below 5 kΩ. Sampling frequency rate was 2048 Hz at an analogue–digital precision of 24 bits. EEGs were recorded in a shielded, sound-attenuated room, where subjects sat in a comfortable armchair. The EEG cap setup took between 20 and 30 min for all sessions—during this time, subjects were sitting in the chair. One week before or after the experiment, a T1-weighted MRI image covering the whole head was acquired. To achieve MRI/EEG coregistration, fiduciary markers were placed at fixed distances from anatomical landmarks identifiable in the participant's anatomical MRIs (tragus, eye center). Brain cortical segmentation was carried out with BrainVoyager (BrainVoyager QX, the Netherlands).

EEG Analysis

The EEG was initially preprocessed with the BESA software (version 5.3, Megis software, Germany). Eye blinkings were corrected using adaptive artifact correction (Ille, Berg, & Scherg, 2002). For the dipole modeling, a four-shell ellipsoidal head model was used with conductivities of 0.33, 0.33, 0.0042, and 1 for the brain, scalp, bone, and CSF, respectively. The period of 0–1500 msec poststimulus (numerical cue, maintenance period; blue shadowed interval in Figure 1) onset was modeled using a standard multiple source model implemented in BESA 5.3 (the FR montage in BESA 5.3), which consists of six dipoles (single sources) modeling the left and right PFC, plus 13 additional regional sources (each regional source consists of three orthogonal dipoles) that were symmetrically placed all around the brain cortex at fixed Talairach coordinates for all subjects (Table 1).

Pattern Classification Preprocessing

The epochs were digitally filtered in the frequency bands of interest: low gamma 30–60, middle gamma 60–100 and high gamma 100–200 Hz using a band-pass zero-shift Butterworth filter (24 dB/octave). The rationale to focus the analysis in the gamma frequency band is that studies in animals suggest that contents of working memory might be contained at >30 Hz oscillations (Pesaran, Pezaris, Sahani, Mitra, & Andersen, 2002). Moreover, in humans, gamma activity shows positive load effects during working memory tasks (Axmacher, Schmitz, Wagner, Elger, & Fell, 2008), and this is accompanied by recent evidence, which suggests that increases of BOLD-fMRI in PFC correlates with increases in >30 Hz oscillations (Michels et al., 2010).

The present study was based on the assumption that during the delay periods each neuron does not have to

Table 1. Talairach Coordinates of the Modeled Cortical Regional Sources

	Talairach Coordinate		
Brain Region	*x*	*y*	*z*
Left PFC	−33	32	25
Right PFC	33	32	25
Medial FpC	0	52	17
Medial FC	0	30	47
Left TAC	−41	5	−23
Right TAC	41	5	−23
Left TPC	−50	−43	−11
Right TPC	−50	−43	−11
Left C	−39	−21	43
Right C	39	−21	43
Medial C	0	−22	61
Left PPC	−33	−70	16
Right PPC	33	−70	16
Medial PC	0	−71	37
Medial Op	0	−87	4

The encoding and maintenance periods were modeled using a standard multiple source model implemented in BESA 5.3 (the FR montage in BESA 5.3), which consists 15 regional sources (each regional source consists of three orthogonal dipoles, thus resulting in 15 × 3 = 45 dipoles) that are symmetrically placed all around the cerebral cortex at fixed Talairach coordinates for all subjects.

FpC = frontopolar cortex; FC = frontocentral cortex; TAC = anterior temporal cortex; TPC = temporo-parietal cortex; C = central cortex; PC = parietal cortex; Op = occipitopolar cortex.

maintain continuous neural activity throughout the delay periods, but working memory is mediated by sequential activation of different neural populations (Baeg et al., 2003). Thus, we hypothesize that such sequential neural activations are reflected as oscillatory patterns which might be captured by analyzing the nonlinear dynamics of EEG recordings. Even though along history linear methods have been developed to study natural systems, it has been shown that they are generally not well described as sums of independent frequencies that can be sensibly decomposed and analyzed as noninteracting and resembled (e.g., Fourier transform). Rather, systematic studies have revealed many natural systems to be more consistent with nonlinear (i.e., state dependent) dynamics, where relationships between state variables cannot be studied independently of the overall system state (Dixon, Milicich, & Sugihara, 1999; Sugihara & May, 1990). Following this concept, Stam and colleagues (1999) suggested that brain rhythms show nonlinear dynamics. Thus, given the nonlinear origin of brain rhythms, the use of nonlinear signal analysis may reveal more information than linear techniques (e.g., power spectra and coherence analysis). These nonlinear methods have been recently used for instance to detect statistical interdependencies between activities in distinct brain areas that are not governed by simple linear functions (Polania, Nitsche, & Paulus, 2011; Stam, Jones, Nolte, Breakspear, & Scheltens, 2007; Stam, Breakspear, van Cappellen van Walsum, & van Dijk, 2003). Following these concepts, the basic assumption of the decoding method used in this study is that the state of the dynamical system (e.g., an epoch of a given dipole) at any given moment may be represented by an embedding vector, where recurrent states are represented by similar embedding vectors; see Takens' (1981) theorem. The preprocessing of the decoding approach used in the present study can be divided into the following steps: (1) Choose a frequency band of interest and band-pass filtering. As previously discussed we select 30–60, 60–100, 100–200 Hz. (2) For a given dipole, construction of the time-delay embedding vectors that represent the dynamical states of the neural system. Let the 1500-msec maintenance period epoch (Figure 1) of a given dipole be considered as a dynamical system X. Next, the time series $x_1, x_2, \ldots, x_N$ is converted into a set of m-dimensional vectors whose components are the time-delayed values of the variables:

$$X_i = (x_i, x_{i+L}, x_{i+2\times L}, x_{i+3\times L}, \ldots, x_{i+(m-1)\times L}),\ X_i \in \mathrm{R}^m, \quad (1)$$

where L is the time lag. Thus, the information in the one-dimensional data was converted to a set of m-dimensional patterns, where $\hat{N} = N - [L * (m - 1)]$ vectors (patterns) can be reconstructed. (3) The state vector must sample the signal at sufficiently short intervals to pick up the fastest oscillation (highest frequency in band-pass filter) and also to be long enough to pick the slowest oscillation (lowest frequency in the band-pass filter). Following Montez, Linkenkaer-Hansen, van Dijk, and Stam (2006) and the Nyquist theorem, a dynamical process must be sampled at minimum twice the highest frequency (HF) of its fluctuation. However, a factor of 3 is often recommended. This yields to the selection of the time lag L in the following way:

$$L = \frac{\mathrm{fs}}{3\mathrm{HF}}, \quad (2)$$

where fs is the sampling frequency in Hz of the EEG recordings. On the other hand, the lowest frequency (LF) has the longest period, therefore, determines the length of the state vector L * (m − 1)

$$L * (m - 1) = \frac{\mathrm{fs}}{\mathrm{LF}} \Longleftrightarrow m = \frac{3 * \mathrm{HF}}{\mathrm{LF}} + 1 \quad (3)$$

(4) Once the dynamical system X is represented in the state space, we define a criterion to test whether a state at different times is similar or recurrent. To this end, we use the autocorrelation integral $C(r, \hat{N})$, which provides information about the likelihood that two randomly chosen patterns will be closer than a cut-off distance r_x:

$$C(r_x, \hat{N}) = \frac{2}{\hat{N}(\hat{N} - w)} \sum_{i=1}^{\hat{N}} \sum_{j=i+w}^{\hat{N}-w} H(r_x - |X_i - X_j|), \quad (4)$$

where $H(\cdot)$ is the Heaviside step function and w is the Theiler correction for autocorrelation (Theiler, 1986). The vertical bars represent the Euclidean distance between the vectors. Regarding the parameter w, Theiler (1986) demonstrated that the nonprevention of the inclusion of states that are similar because of autocorrelation effects can lead to spurious estimates of dimension, that is, the vectors starting inside the w window are likely not to represent a recurrence of the reference state but the state itself. Thus, w is chosen to be at least twice the length of the embedding vector. (5) The next step is to find r_x by setting an autocorrelation reference $P_{\mathrm{ref}} = C(\hat{N}, r_x)$, which in this study is set to $P_{\mathrm{ref}} = 0.5$. Notice that having P_{ref} constant means that the number of recurrences is fixed, whether they exist or not. Once having r_x at $P_{\mathrm{ref}} = 0.5$, Equation 4 can be seen as vector B with $\hat{N}(\hat{N} - w)$ binary entries

$$b_{ij} = \begin{cases} 1 & \text{if } H(r_x - |X_i - X_j|) = 1 \\ 0 & \text{if } H(r_x - |X_i - X_j|) = 0 \end{cases} \quad (5)$$

where $H(\cdot)$ is the Heaviside step function in Equation 2. Hence, notice that at $P_{\mathrm{ref}} = 0.5$ we will have 50% of the entries in B being 1. The reason to choose this value for P_{ref} was that we wanted to have a balanced number of entries being either 1 or 0 in the binary vector B, which gives results in a bell-shaped distribution centered at half

of the number of trials. The information in B is afterwards used in the binary pattern classifier described in the following subsection.

Pattern Classification

When a histogram from a set of binary vectors B belonging to a given stimulus (say cued letter L trials) is plotted, a bell-shaped distribution centered at the half of number of trials is generated. This is expected because we set $P_{\text{ref}} = 0.5$ (e.g., if we would have set $P_{\text{ref}} = 0.25$, we would expect the distribution to be centered at 25% of the number of trials). Let us consider that we want to discriminate whether a subject retains the cued letter L or T. First from the 240 trials of each experiment, a subset of 180 training trials (three out of the four experimental blocks) is used to build a histogram for cued letter T stimuli

$$S_{\text{T}} = \sum_{n}^{N_{\text{T}}} B_n \tag{6}$$

and a second histogram for cued letter L stimuli

$$S_{\text{L}} = \sum_{n}^{N_{\text{L}}} B_n \tag{7}$$

where N_{L} and N_{T} are the number of trials for the cued letter L and letter T stimuli, respectively, and B_n is the binary vector of trial n. Thus, S_{L} is a vector of size $\hat{N}(\hat{N} - w)$, whose entries $S_{\text{L}(ij)}$ contain information of the number of trials where b_{ij} is 1 for the set of L stimuli (the same holds for the S_{T}). The second step is the following: we ask for which entries ij in S_{L} and S_{T} the following conditions hold

$$\text{Condition 1} : (S_{\text{L}(ij)} > (T * N_{\text{L}})) \wedge (S_{\text{T}(ij)} < (0.5 * N_{\text{T}})) \tag{8}$$

or

$$\text{Condition 2} : (S_{\text{T}(ij)} > (T * N_{\text{T}})) \wedge (S_{\text{L}(ij)} < (0.5 * N_{\text{L}})) \tag{9}$$

that is, if we set the threshold T to $T = 0.6$, in Condition 1 we are asking for which entries b_{ij} is 1 for more than 60% of the letter L training trials, conjoined ($\wedge$) with the condition that b_{ij} is 1 for less than 50% of the cued letter T training trials. Thus, if for a given dipole there are a total of d entries ij where the conditions from Equations 8 and 9 hold, then these d binary features are used as a pattern for each single trial. In the present study we looked for a threshold T where $d = 1, 2, \ldots, 500$. Notice that the binary vector B has $\hat{N}(\hat{N} - w)$ binary entries; however, by adjusting T we restrict the number of classification features d to 1 up to 500 (which is around two times the total number of trials in one experiment) to avoid circularity bias (Dosenbach et al., 2010; Pereira, Mitchell, & Botvinick, 2009); notice that the size of B for a 1500-msec epoch at a sampling frequency of 2 kHz is $\hat{N}(\hat{N} - w) >> 10^3$ (see Equation 2), and the number of characters presented in each experiment is $60 \times 4 = 240$. Once we have selected the classification features (d), consider that we let ω_1 be the category letter L, ω_2 be the category letter T, and for each trial we let $\widetilde{\text{X}} = (\tilde{x}_1, \ldots, \tilde{x}_d)^t$, where the components $\tilde{x}_i$ are either 0 or 1, with probabilities

$$p_i = \Pr[\tilde{x}_i = 1 | \omega_1] \tag{10}$$

and

$$q_i = \Pr[\tilde{x}_i = 1 | \omega_2], \tag{11}$$

where $Pr[\cdot]$ is the probability of a condition being met. This classification model gives a yes/no answer about the pattern. If $p_i > q_i$, the ith feature is expected to give a ''yes'' answer more frequently when the category is ω_1 than when it is ω_2. As it is commonly done in pattern classification literature we write $P(\widetilde{\text{X}} | \omega_i)$ as the product of the probabilities for the components of $\widetilde{\text{X}}$:

$$P(\widetilde{\text{X}} | \omega_1) = \prod_{i=1}^{d} p_i^{\tilde{x}_i} (1 - p_i)^{1 - \tilde{x}_i} \tag{12}$$

and

$$P(\widetilde{\text{X}} | \omega_2) = \prod_{i=1}^{d} q_i^{\tilde{x}_i} (1 - q_i)^{1 - \tilde{x}_i}, \tag{13}$$

However, it is important to notice that conditional independence cannot be assumed given that the elements of $\widetilde{\text{X}}$ are derived from the same time series. Thus, the calculations in the present study were simplified as they are written in Equations 12 and 13.

The likelihood ratio of Equations 12 and 13 is given by

$$\frac{P(\widetilde{\text{X}} | \omega_1)}{P(\widetilde{\text{X}} | \omega_2)} = \prod_{i=1}^{d} \left(\frac{p_i}{q_i} \right)^{\tilde{x}_i} \left(\frac{1 - p_i}{1 - q_i} \right)^{1 - \tilde{x}_i} \tag{14}$$

Let $g(\widetilde{\text{X}})$ be a discriminant function for a system with continuous features as suggested by Duda, Hart, and Stork (2001),

$$g(\widetilde{\text{X}}) = \ln \left(\frac{p(\widetilde{\text{X}} | \omega_1)}{p(\widetilde{\text{X}} | \omega_2)} \right) + \ln \left(\frac{P(\omega_1)}{P(\omega_2)} \right) \tag{15}$$

Then replacing terms of Equation 14 in Equation 15 yields to the following discriminant function for binary features,

$$g(\widetilde{\text{X}}) = \sum_{i=1}^{d} \left[\tilde{x}_i \ln \frac{p_i}{q_i} + (1 - \tilde{x}_i) \ln \frac{1 - p_i}{1 - q_i} \right] + \ln \frac{P(\omega_1)}{P(\omega_2)} \tag{16}$$

Notice that the discriminant function $g(X)$ is linear in x_i; thus, we can write,

$$g(\tilde{\mathrm{X}}) = \sum_{i=1}^{d} a_i \tilde{x}_i + b_0 \tag{17}$$

where

$$a_i = \ln \frac{p_i(1-q_i)}{q_i(1-p_i)} \; i = 1,\ldots,d \tag{18}$$

and

$$b_0 = \sum_{i=1}^{d} \ln \frac{1-p_i}{1-q_i} + \ln \frac{P(\omega_1)}{P(\omega_2)} \tag{19}$$

Once the discriminant function has been generated, we use the patterns from the test block to decide for a given pattern: ω_1 if $g(\tilde{\mathrm{X}}) > 0$ and ω_2 if $g(\tilde{\mathrm{X}}) < 0$. In our experiment each character has the same probability to appear, thus $P(\omega_1) = P(\omega_1) = 0.5$. The fourfold cross validation was achieved by repeating this procedure independently, with each block acting as a test dataset once, while the other three blocks were used as training data sets.

Statistical Analysis Classification

At the individual level for each dipole, we do a bootstrap-*t* estimation of *p* that result from the binomial test (i.e., binomial test is the function to be bootstrapped and the classification hits/misses are the vectors to be shuffled; the hypothesized probability of success in the binomial test was one-sided and set to >50%) and see if the hypothesized $p = .05$ is in the confidence interval. At the group level for each dipole, we calculate the bootstrap-*t* confidence intervals using the accuracies of each individual. Then, for an alpha = 0.05/95% confidence interval, we see if the lower boundary of the confidence interval is >50%. The number of bootstrap samples used to estimate the distribution of the bootstrap-*t* statistic was 2000. Significance level was corrected for multiple comparisons by setting an alpha (Bonferroni) = 0.05/95%, and then we see whether lower boundary of the confidence interval is >50%.

RESULTS

Behavioral Results

Behavioral data showed that RTs were not different when subjects had to remember letter T and L ($t(11) = 0.09$, $p > .05$). When subjects had to remember the first cued letter, RTs were slightly prolonged, compared to the RTs when subjects had to remember the second cued letter ($t(11) = 0.4, p > .05$). Subjects showed equally good performance when letter T or L had to be remembered ($t(11) = 0.81, p > .05$) and also when the first or second cued letter had to be remembered ($t(11) = 1.42, p > .05$) (Figure 2). Additionally using the information of the behavioral data, we calculated the percentage of appearances of the letter T as first stimulus, obtaining the following results: all subjects interval, 48–54%; mean = 50.9, *SD* = 2.1. Afterwards, we computed the number of times that ''T'' was the first stimulus and the cue was ''1'' in the maintenance period: all subjects interval, 44–54%; mean = 49.1, *SD* = 3.2. Then, we computed the number of times that ''L'' was the first stimulus and the cue was ''1'' in the maintenance period: All subjects interval, 44–55%; mean = 51, *SD* = 2.5. For all subjects and all of the cases described above, the *p* value of a binomial test was always $p > .05$. These results suggest that numeric cues during maintenance periods were equally associated to each letter and their order of appearance at the encoding stage was also randomized properly.

Classification

After classification was carried out across all modeled cortical sources, the activity patterns from left PFC were highly predictive of the letter held in working memory, with prediction accuracies of ~62% in the 30–60 Hz range, ~73% in the 60–100 Hz range, and ~82% in the 100–200 Hz range (Figure 3A). For 60–100 and 100–200 Hz ranges, decoding accuracy clearly exceeded the chance level of correct letter identification (lower boundary of the confidence interval in the bootstrap-*t* statistic corrected for multiple comparisons was >50%) when the number of classification features *d* was $200 < d < 500$ (*d* is the number of classification features obtained from the binary arrays after calculating the autocorrelation integral; see Methods). At these frequency bands, decoding accuracy also proved to be highly reliable at the individual subject level (for all subjects $p = .05$ was in the confidence interval after binomial test was bootstrapped; see Methods). Beyond the left PFC, left PPC also showed high accuracy, but only in the 100–200 Hz frequency range (Figure 3B) (accuracy ~62%, significant for noncorrected alpha; however, it did not survive Bonferroni correction). At the individual level, for 9 of 12 subjects, the hypothesized $p = .05$ was in the confidence interval after binomial test was bootstrapped. For the rest of the modeled cortical sources, decoding accuracy was at chance level.

In a second set of calculations, we investigated whether PFC also temporarily stores working memory information during the encoding periods. Thus, we repeated the calculations, this time analyzing the oscillatory patterns during encoding periods (encode_1 and encode_2 periods in Figure 1). Decoding accuracy in the left PFC was highly reliable at 60–100 and 100–200 Hz frequency bands (lower boundary of the confidence interval in the bootstrap-*t* statistic corrected for multiple comparisons was >50%). For the rest of modeled cortical sources, decoding accuracy was at chance level. Thus, results are indicative for a participation of PFC also in the active storage of information during encoding periods (Figure 4).

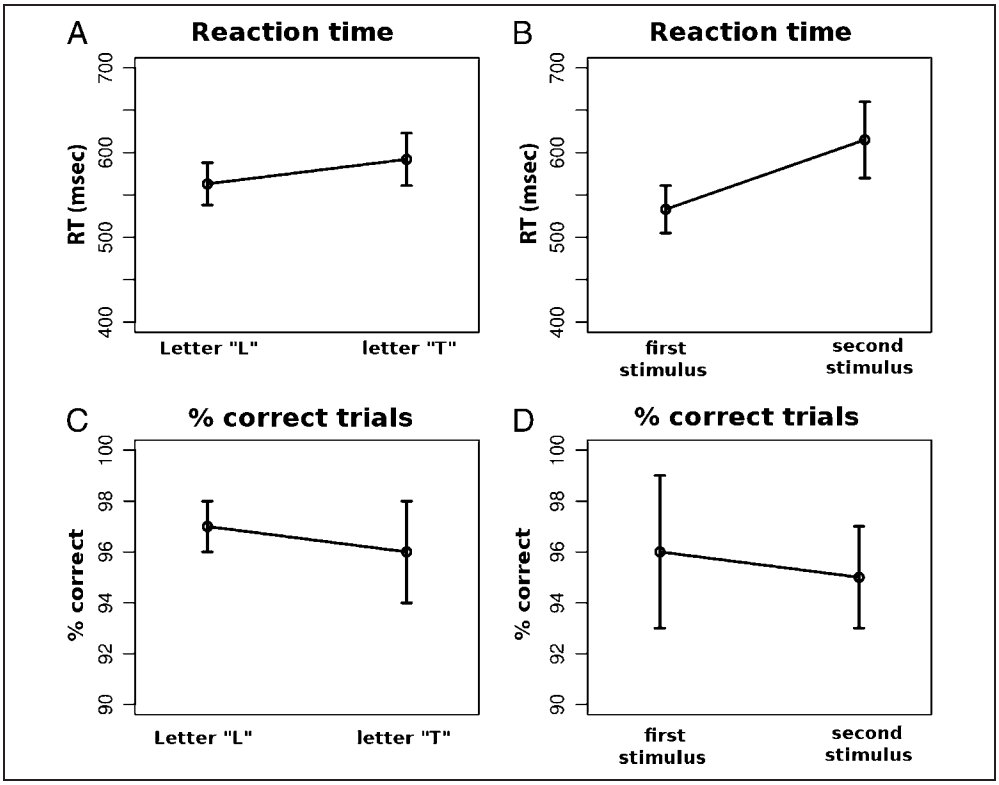

Figure 2. Behavioral results. (A) RTs were not different when subjects had to remember letter T and L. (B) When subjects had to remember the first cued letter, RTs were slightly prolonged, compared with the RTs when subjects had to remember the second cued letter. (C) Subjects showed equally good performance when the letter T or L had to be remembered and (D) also when the first or second cued letter had to be remembered. Paired *t* tests, $p > .05$ for A–D. Error bars represent *SEM*.

As a control set of calculations, we determined whether systematic eye movements could account for our EEG decoding approach (Yuval-Greenberg, Tomer, Keren, Nelken, & Deouell, 2008). Thus, we performed our calculations on the VEOG and HEOG recordings. Additionally, since we performed our analysis in the cortical source space, we modeled dipoles in the eye orbits of each individual, in a similar way as studied by Yuval-Greenberg and colleagues (2008), and our decoding approach was applied over each eye orbit dipole. The accuracies resulting from the decoding analysis on the EOGs, and the dipoles modeled in the eye orbits were clearly at chance level (lower boundary of the bootstrapping-*t* statistic confidence interval was <50% even for an uncorrected alpha = 0.05). Thus,

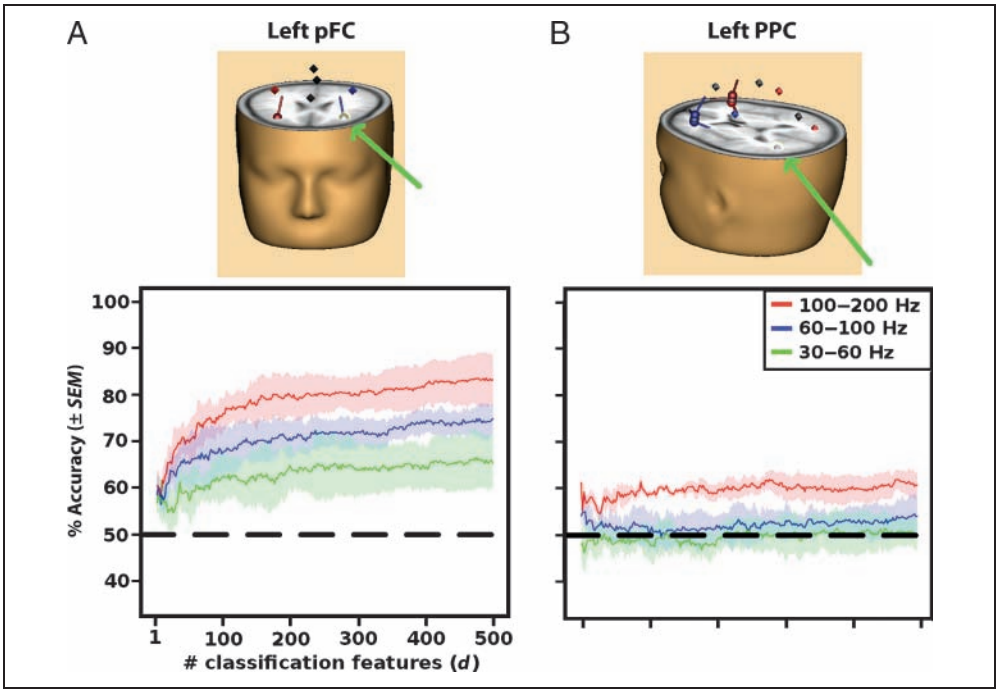

Figure 3. Classification performances maintenance period. Shown are the classifier performances in the left PFC (A) and the left PPC (B) in the 30–60 (green curve), 60–100 (blue curve), and 100–200 Hz (red curve) frequency bands. The *x* axis shows the number of classification features *d*. Dashed lines show chance level (50%). Shadowed regions represent the *SEM*. Green arrows indicate the location of the respective modeled source.

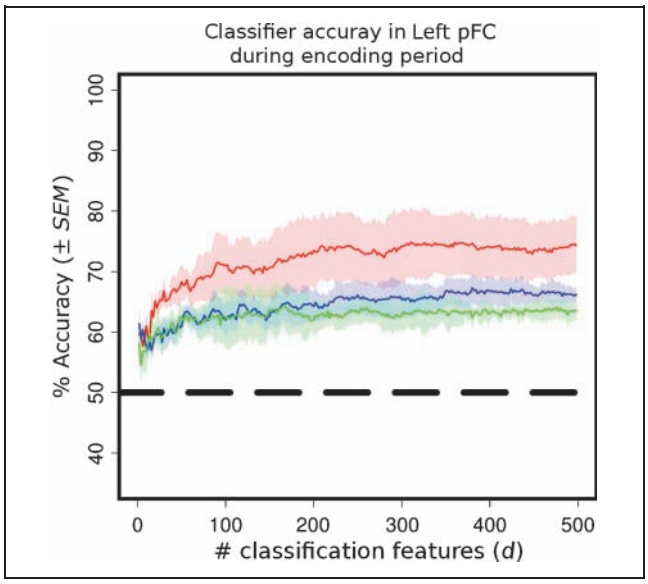

Figure 4. Decoding accuracies in the left PFC during encoding period. Shown are the decoding accuracies in the cortical source modeled in the left PFC during the encoding period in the 30–60 (green curve), 60–100 (blue curve), and 100–200 Hz (red curve) frequency bands. Dashed line shows chance level (50%). Shadowed regions represent the *SEM*.

possible eye movement activity added on EEG recordings was not a reliable predictor of the letter held in working memory.

Exploratory Cross-frequency Analysis

In an exploratory analysis, we investigated whether in the left PFC theta frequency band (4–8 Hz) was phase locked to the instantaneous amplitude of the gamma oscillations. Several simulation and electrophysiological studies have demonstrated this mechanism as a candidate for establishing memory functions in the brain (Sauseng et al., 2009; Jensen & Colgin, 2007; Canolty et al., 2006; Fell et al., 2003). Based on this evidence, we calculated the theta-phase gamma-amplitude modulation index (*M_index*) following the procedure proposed by Canolty et al. (2006). In brief, single trials $x(t)$ (starting when the first stimulus is presented and finishing at subject probe response) are filtered into low-frequency bands (with center frequencies 2–20 Hz, in 1 Hz steps with 1 Hz bandwidths) and high-frequency bands (with center frequencies 60–200 Hz, in 5 Hz steps with 4 Hz bandwidths), generating $x_{fP}(t)$ and $x_{fA}(t)$, respectively. Using the Hilbert transform, the analytic phase $\phi(t)$ and $A(t)$ amplitude were obtained from $x_{fP}(t)$ and $x_{fA}(t)$ respectively to form the composite raw signal $z(t) = A(t)e^{i\phi(t)}$, and the raw modulation index was computed as $M_{\text{raw}} = |\langle z(t) \rangle|$. The length (or modulus) of M_{raw}, compared with the distribution of surrogate lengths provides a normalized z-scored modulation index $M_{\text{index}} = (M_{\text{raw}} - \mu)/\sigma$, where μ is the mean of 500 surrogate lengths and σ their standard deviation. The M_{index} was averaged over trials for each subject. The left PFC dipole (Figure 3A) shows a high modulation index ($z > 2$) for the analytic phase, 4–7 Hz, and analytic amplitude, >100 Hz (Figure 5A). Afterwards, instantaneous phase at 6 Hz as well as instantaneous amplitude values (each normalized within trials) for the 100–200 Hz frequency band were concatenated for all trials. Next, the amplitude values of 100–200 Hz were sorted according to instantaneous 6-Hz phase values. Amplitude values of 100–200 Hz were then averaged for 80 segments of $2\pi/80$ in respect to the sorted 6- Hz cycle. A similar

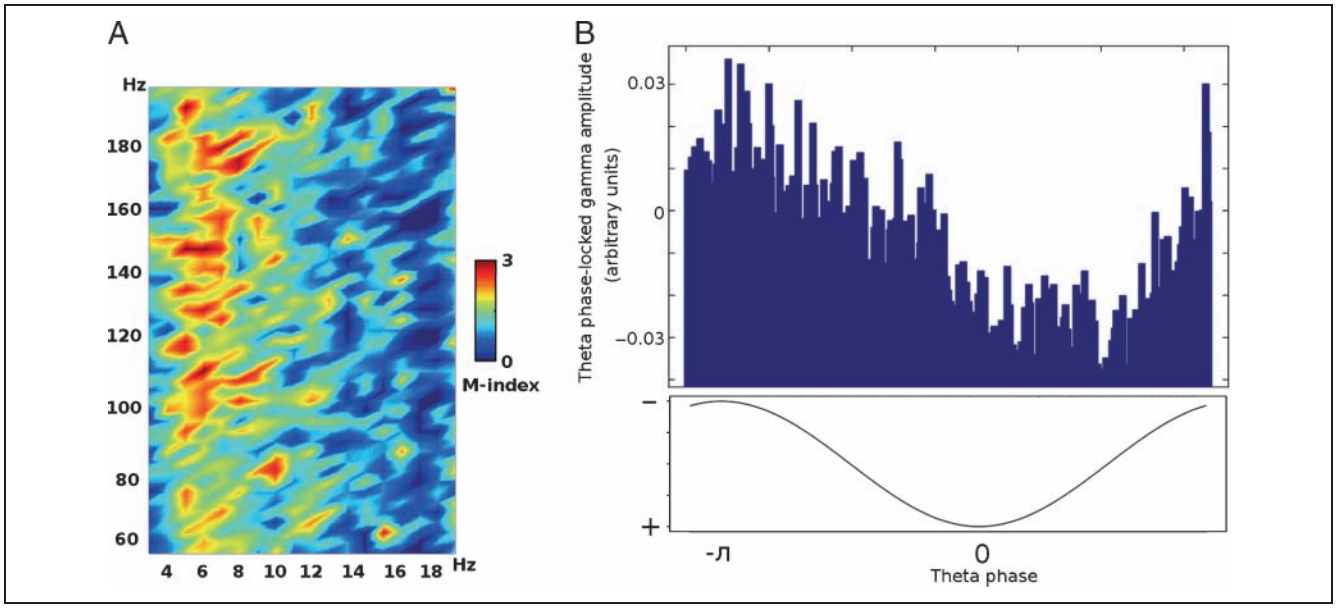

Figure 5. Cross-frequency analysis. (A) Shown is the modulation index as a function of the analytic amplitude (60–200 Hz) and analytic phase (3–20 Hz) for the left PFC dipole. Maximal coupling was 4–7 Hz and >100 Hz. (B) Instantaneous amplitudes for frequencies >100 Hz were sorted according to the instantaneous theta 6 Hz phase. Increased >100 Hz amplitude is phase-locked to the negative peak of a theta cycle.

approach to study low-frequency phase-dependant high-frequency modulation was proposed previously (Sauseng et al., 2009; Demiralp et al., 2007). Similar to previous studies (Sauseng et al., 2009; Canolty et al., 2006), increased high-gamma amplitude, is locked to the negative peak of the theta cycle (Figure 5B). This is confirmed by a significant main effect of theta phase angle with high-gamma amplitude as dependent variable by ANOVA: $F_{79/869} = 7.95$, $p < .0001$.

DISCUSSION

The results of the study provide evidence that PFC is involved in the active storage of information during working memory tasks in the human brain. Additionally, they show that the contents of visual working memory are preferentially stored within high-gamma oscillatory patterns and, moreover, such patterns can be non-invasively decoded.

The present finding is in principle accordance with animal studies, where it was shown via invasive recordings that specific information about the contents of working memory can be decoded from population activity in PFC (Averbeck & Lee, 2007; Baeg et al., 2003). These studies suggested that working memory is robustly represented in PFC cortex, and this representation is mediated by the sequential activation of neural populations. In line with these results, a recent computational–experimental study provides evidence for higher-order precise spike synchrony in PFC during visual working memory (Pipa & Munk, 2011). The investigators show with multiple simultaneously recording units in the ventral PFC of nonhuman primates that neurons participate in 3-msec precise synchronous discharges distributed across multiple sites. Interestingly, the authors provide evidence that such synchronous firing is specific for the memorized visual stimuli. Taking together, we hypothesized that this sequential activation of local neuronal assemblies encode the stimulus representations during the delay periods, and also that this sequential activation is reflected in EEG scalp recordings as oscillatory patterns in the gamma frequency band. The reason to believe that the stimulus held in memory is represented in high-frequency oscillatory activity is that studies in animals suggest that contents of working memory might be contained at >30 Hz oscillations (Pesaran et al., 2002). Moreover, in humans gamma activity shows positive load effects during working memory tasks (Axmacher et al., 2008); and this is accompanied by recent evidence which suggests that increases of BOLD-fMRI in PFC correlates with increases in >30 Hz oscillations (Michels et al., 2010). However, we still had the problem of how to identify and differentiate such high-frequency oscillatory patterns from scalp recordings. About a decade ago, Stam and colleagues (1999) suggested that brain rhythms show nonlinear dynamics. Thus, given the nonlinear origin of brain rhythms, the use of nonlinear signal analysis reveals more information than linear techniques (e.g., power spectra and coherence analysis). Hence, we intended to extract and differentiate oscillatory activity patterns by means of nonlinear signal analysis. In line with our hypotheses, we were able to decode with relatively high accuracy the contents of visual working memory from high-gamma oscillatory patterns not only from PFC, but also in the PPC, which have been regions consistently shown to participate working memory (Axmacher et al., 2008; Miller, Deouell, Dam, Knight, & D'Esposito, 2008; D'Esposito, 2007; Sakai & Passingham, 2003). In accordance with our results, it was demonstrated that gamma oscillations present a temporal structure in the parietal cortex of monkeys during working memory tasks (Pesaran et al., 2002). Based on that temporal structure, the authors were able to decode from local field potentials in a single-trial basis the contents of working memory. In our results, it should be noticed, however, that decoding accuracies in the 60–100 and 100–200 Hz frequency bands were always higher in PFC compared with the rest of modeled dipoles including the left PPC (paired t tests, $p < .001$ for all d). One reason for this might be that neuronal spiking pattern is more ''bursty'' in the PFC relative to more posterior cortical areas (Shinomoto et al., 2009), which might be beneficial for the nonlinear analysis used in the present study (notice that we split maintenance epochs in small vectors [patterns]; see Methods). Thus the fact that in this study decoding accuracies in PPC were much lower than in PFC does not necessarily imply that more posterior regions do not participate in the active storage during delay periods (Axmacher et al., 2008).

Our results are consistent with the notion that theta oscillations organize local functional assemblies, defined by faster oscillations, which encode object representations (Fell & Axmacher, 2011; Axmacher et al., 2010). In line with this idea, a recent study in nonhuman primates provides evidence that high-gamma oscillations (>60 Hz) in the PFC—which are phase-locked to lower frequency oscillations—carries information about the visual stimuli during the delay period in a working memory task, while low-frequency oscillations were correlated with performance but not stimulus content (Pipa et al., 2009). Thus, cross-frequency synchronization might constitute an important part of the maintenance process and provide a link of the stimuli representations between distinct cortical regions. This idea is in accordance with physiological data, which suggest that each brain region, although forming part of the working memory functional network, contributes to the active maintenance by the nature of the representations that are coded within each region (D'Esposito, 2007; McIntosh, 2000). Here, we speculate that another possibility for PFC to present higher levels of accuracy in our results is that PFC integrates the representations that are being locally processed in more detail in anterior regions. Interestingly, we found left PFC and PPC (but not the right) to present higher levels of accuracy, which is in line with previous studies that have demonstrated that left PFC/PPC (and not right) are activated during working memory maintenance in letter recognition paradigms (Oztekin, McElree, Staresina, & Davachi, 2009). Nevertheless, we do

not rule out that, in addition to the role of PFC in the temporal storage of stimuli, it also exerts top–down control through theta–gamma coupling (Canolty et al., 2006) to keep neural representations of behaviorally relevant sensory information activated in more anterior regions activated (Miller & D'Esposito, 2005; Petrides, 2005).

Taking together, our results suggest that synchronous firing (~3–10 msec) in PFC during delay periods of working memory tasks, which is specific for the memorized stimulus (Pipa & Munk, 2011), can be captured as oscillatory patterns by means of EEG recordings and subsequently decoded in the human brain. The results of our study may further support the hypothesis of sensory coding of feature conjunctions as stated in the binding hypothesis. As a concluding remark, the ability to noninvasively decode the contents of working memory during encoding and maintenance periods is promising in applications such as brain computer interfaces (Santhanam, Ryu, Yu, Afshar, & Shenoy, 2006). Additionally, the presented methodology might be useful in computation of value function during planning and decision-making processes (Curtis & Lee, 2010). Decoding of the short-term storage of behaviorally relevant information might be useful to understand how we make predictions about the future and how we select behaviors that are likely to result in the most beneficial outcomes (Seo, Barraclough, & Lee, 2009; Seo & Lee, 2009).

Acknowledgments

This work was supported by the Rose Foundation. We thank J. D. Florez for helpful discussions.

Reprint requests should be sent to Rafael Polanía, Department of Clinical Neurophysiology, Georg-August University of Göttingen, Robert Koch Str 40, 37075 Göttingen, Germany, or via e-mail: rafael.polania@med.uni-goettingen.de.

REFERENCES

Averbeck, B. B., & Lee, D. (2007). Prefrontal neural correlates of memory for sequences. *Journal of Neuroscience, 27,* 2204–2211.

Axmacher, N., Henseler, M. M., Jensen, O., Weinreich, I., Elger, C. E., & Fell, J. (2010). Cross-frequency coupling supports multi-item working memory in the human hippocampus. *Proceedings of the National Academy of Sciences, U.S.A., 107,* 3228–3233.

Axmacher, N., Schmitz, D. P., Wagner, T., Elger, C. E., & Fell, J. (2008). Interactions between medial temporal lobe, prefrontal cortex, and inferior temporal regions during visual working memory: A combined intracranial EEG and functional magnetic resonance imaging study. *Journal of Neuroscience, 28,* 7304–7312.

Baeg, E. H., Kim, Y. B., Huh, K., Mook-Jung, I., Kim, H. T., & Jung, M. W. (2003). Dynamics of population code for working memory in the prefrontal cortex. *Neuron, 40,* 177–188.

Brunel, N., & Wang, X. J. (2001). Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition. *Journal of Computational Neuroscience, 11,* 63–85.

Canolty, R. T., Edwards, E., Dalal, S. S., Soltani, M., Nagarajan, S. S., Kirsch, H. E., et al. (2006). High gamma power is phase-locked to theta oscillations in human neocortex. *Science, 313,* 1626–1628.

Champod, A. S., & Petrides, M. (2007). Dissociable roles of the posterior parietal and the prefrontal cortex in manipulation and monitoring processes. *Proceedings of the National Academy of Sciences, U.S.A., 104,* 14837–14842.

Champod, A. S., & Petrides, M. (2010). Dissociation within the frontoparietal network in verbal working memory: A parametric functional magnetic resonance imaging study. *Journal of Neuroscience, 30,* 3849–3856.

Cohen, J. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., et al. (1997). Temporal dynamics of brain activation during a working memory task. *Nature, 386,* 604–608.

Curtis, C. E., & D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends in Cognitive Sciences, 7,* 415–423.

Curtis, C. E., & Lee, D. (2010). Beyond working memory: The role of persistent activity in decision making. *Trends in Cognitive Sciences, 14,* 216–222.

Curtis, C. E., Rao, V. Y., & D'Esposito, M. (2004). Maintenance of spatial and motor codes during oculomotor delayed response tasks. *Journal of Neuroscience, 24,* 3944–3952.

Demiralp, T., Bayraktaroglu, Z., Lenz, D., Junge, S., Busch, N. A., Maess, B., et al. (2007). Gamma amplitudes are coupled to theta phase in human EEG during visual perception. *International Journal of Psychophysiology, 64,* 24–30.

D'Esposito, M. (2007). From cognitive to neural models of working memory. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 362,* 761–772.

Dixon, P. A., Milicich, M. J., & Sugihara, G. (1999). Episodic fluctuations in larval supply. *Science, 283,* 1528–1530.

Dosenbach, N. U., Nardos, B., Cohen, A. L., Fair, D. A., Power, J. D., Church, J. A., et al. (2010). Prediction of individual brain maturity using fMRI. *Science, 329,* 1358–1361.

Duda, O. R., Hart, P. E., & Stork, D. G. (2001). *Pattern classification*. New York: Wiley.

Fell, J., & Axmacher, N. (2011). The role of phase synchronization in memory processes. *Nature Reviews Neuroscience, 12,* 105–118.

Fell, J., Klaver, P., Elfadil, H., Schaller, C., Elger, C. E., & Fernandez, G. (2003). Rhinal-hippocampal theta coherence during declarative memory formation: Interaction with gamma synchronization? *European Journal of Neuroscience, 17,* 1082–1088.

Fuentemilla, L., Penny, W. D., Cashdollar, N., Bunzeck, N., & Duzel, E. (2010). Theta-coupled periodic replay in working memory. *Current Biology, 20,* 606–612.

Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature, 458,* 632–635.

Ille, N., Berg, P., & Scherg, M. (2002). Artifact correction of the ongoing EEG using spatial filters based on artifact and brain signal topographies. *Journal of Clinical Neurophysiology, 19,* 113–124.

Jensen, O., & Colgin, L. L. (2007). Cross-frequency coupling between neuronal oscillations. *Trends in Cognitive Sciences, 11,* 267–269.

Jensen, O., & Lisman, J. E. (2005). Hippocampal sequence-encoding driven by a cortical multi-item working memory buffer. *Trends in Neurosciences, 28,* 67–72.

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience, 8,* 679–685.

Lisman, J. E. (1999). Relating hippocampal circuitry to function: Recall of memory sequences by reciprocal dentate-CA3 interactions. *Neuron, 22,* 233–242.
McIntosh, A. R. (2000). Towards a network theory of cognition. *Neural Networks, 13,* 861–870.
Michels, L., Bucher, K., Luchinger, R., Klaver, P., Martin, E., Jeanmonod, D., et al. (2010). Simultaneous EEG-fMRI during a working memory task: Modulations in low and high frequency bands. *PLoS One, 5,* e10298.
Miller, B. T., Deouell, L. Y., Dam, C., Knight, R. T., & D'Esposito, M. (2008). Spatio-temporal dynamics of neural mechanisms underlying component operations in working memory. *Brain Research, 1206,* 61–75.
Miller, B. T., & D'Esposito, M. (2005). Searching for ''the top'' in top–down control. *Neuron, 48,* 535–538.
Montez, T., Linkenkaer-Hansen, K., van Dijk, B. W., & Stam, C. J. (2006). Synchronization likelihood with explicit time-frequency priors. *Neuroimage, 33,* 1117–1125.
Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9,* 97–113.
Oztekin, I., McElree, B., Staresina, B. P., & Davachi, L. (2009). Working memory retrieval: Contributions of the left prefrontal cortex, the left posterior parietal cortex, and the hippocampus. *Journal of Cognitive Neuroscience, 21,* 581–593.
Pereira, F., Mitchell, T., & Botvinick, M. (2009). Machine learning classifiers and fMRI: A tutorial overview. *Neuroimage, 45(1 Suppl.),* S199–S209.
Pesaran, B., Pezaris, J. S., Sahani, M., Mitra, P. P., & Andersen, R. A. (2002). Temporal structure in neuronal activity during working memory in macaque parietal cortex. *Nature Neuroscience, 5,* 805–811.
Petrides, M. (2000). The role of the mid-dorsolateral prefrontal cortex in working memory. *Experimental Brain Research, 133,* 44–54.
Petrides, M. (2005). Lateral prefrontal cortex: Architectonic and functional organization. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 360,* 781–795.
Pipa, G., & Munk, M. H. (2011). Higher order spike synchrony in prefrontal cortex during visual memory. *Frontiers in Computational Neuroscience, 5,* 23.
Pipa, G., Stadtler, E. S., Rodriguez, E. F., Waltz, J. A., Muckli, L. F., Singer, W., et al. (2009). Performance- and stimulus-dependent oscillations in monkey prefrontal cortex during short-term memory. *Frontiers in Integrative Neuroscience, 3,* 25.
Polania, R., Nitsche, M. A., & Paulus, W. (2011). Modulating functional connectivity patterns and topological functional organization of the human brain with transcranial direct current stimulation. *Human Brain Mapping, 32,* 1236–1249.
Sakai, K., & Passingham, R. E. (2003). Prefrontal interactions reflect future task operations. *Nature Neuroscience, 6,* 75–81.
Santhanam, G., Ryu, S. I., Yu, B. M., Afshar, A., & Shenoy, K. V. (2006). A high-performance brain-computer interface. *Nature, 442,* 195–198.
Sauseng, P., Klimesch, W., Heise, K. F., Gruber, W. R., Holz, E., Karim, A. A., et al. (2009). Brain oscillatory substrates of visual short-term memory capacity. *Current Biology, 19,* 1846–1852.
Seo, H., Barraclough, D. J., & Lee, D. (2009). Lateral intraparietal cortex and reinforcement learning during a mixed-strategy game. *Journal of Neuroscience, 29,* 7278–7289.
Seo, H., & Lee, D. (2009). Neuroscience: Persistent feedback. *Nature, 461,* 50–51.
Shinomoto, S., Kim, H., Shimokawa, T., Matsuno, N., Funahashi, S., Shima, K., et al. (2009). Relating neuronal firing patterns to functional differentiation of cerebral cortex. *PLoS Computational Biology, 5,* e1000433.
Stam, C. J., Breakspear, M., van Cappellen van Walsum, A. M., & van Dijk, B. W. (2003). Nonlinear synchronization in EEG and whole-head MEG recordings of healthy subjects. *Human Brain Mapping, 19,* 63–78.
Stam, C. J., Jones, B. F., Nolte, G., Breakspear, M., & Scheltens, P. (2007). Small-world networks and functional connectivity in Alzheimer's disease. *Cerebral Cortex, 17,* 92–99.
Stam, C. J., Pijn, J. P., Suffczynski, P., & Lopes da Silva, F. H. (1999). Dynamics of the human alpha rhythm: Evidence for non-linearity? *Clinical Neurophysiology, 110,* 1801–1813.
Sugihara, G., & May, R. M. (1990). Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature, 344,* 734–741.
Takens, F. (1981). Detecting strange attractors in turbulence. *Lecture Notes in Mathematics, 898,* 366–381.
Theiler, J. (1986). Spurious dimension from correlation algorithms applied to limited time-series data. *Physical Review A, 34,* 2427–2432.
Wang, X. J. (2001). Synaptic reverberation underlying mnemonic persistent activity. *Trends in Neuroscience, 24,* 455–463.
Yuval-Greenberg, S., Tomer, O., Keren, A. S., Nelken, I., & Deouell, L. Y. (2008). Transient induced gamma-band response in EEG as a manifestation of miniature saccades. *Neuron, 58,* 429–441.