Title:

Dynamic Predictions: Oscillatory Mechanisms Underlying Multisensory Sequence Processing

Running title: Multisensory Sequence Prediction

Peng Wang[a,*], Alexander Maye[a], Jonathan Daume[a], Gui Xue[b] and Andreas K. Engel[a]

[a] Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, Martinistr. 52, 20246 Hamburg, Germany

[b] State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, 100875, Beijing, China

[*] *Correspondence address:*

Peng Wang, Phone: +49-40-741054680, Fax: +49-40-741057126, E-mail: p.wang@uke.de

Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, Martinistr. 52, 20246 Hamburg, Germany.

Abstract

Neural oscillations have been proposed to be involved in predictive processing by frequency-specific modulation of either power or phase. While this is supported by substantial evidence on unimodal processing, only few studies are currently available that have addressed the role of oscillatory activity in multisensory predictions. In the present study, we have recorded MEG during a serial prediction task in which participants had to acquire stimulus sequences and to monitor whether subsequent probe items complied with the sequence. The sequences comprised combinations of visual and auditory stimuli, but on a given trial, only one of the modalities was task-relevant and the other had to be ignored. Task-related changes of power and coupling of neural oscillations were analyzed using a data-driven clustering approach. We observed prediction-related theta-band coupling in a network involving cingulate, premotor, prefrontal and superior temporal regions. Behavioral performance of the participants correlated with phase delays between remote regions in this network. Two additional clusters were detected where beta power were stronger when the previous or next stimuli were same as the current one in the sequence. These clusters interacted with the theta-band prediction network through cross-frequency coupling. Furthermore, we observed asymmetrical effects of attention to the visual and auditory modalities in a network distinguished by changes in alpha power, where alpha phases modulated response times. Overall, our results provide novel insights into the neural dynamics underlying multisensory predictions and suggest that oscillations in multiple frequency ranges as well as coupling within and across frequencies may be critical for sequence processing.

Keywords

Attention, imaginary part of coherency, multisensory processing, phase-amplitude coupling.

Introduction

While natural events and objects are often jointly detected by more than one sensory system, the perceptual and behavioral consequences of which are often not just the sum of those evoked by each sensory component alone. However, surprisingly little is known regarding the mechanisms underlying crossmodal interaction in the human brain. Only few studies have so far addressed the issue of dynamic neural interactions in crossmodal processing (Senkowski et al. 2008; Hipp et al. 2011; Senkowski et al. 2014; Wang et al. 2019). Recent EEG studies have shown that integration of visual and auditory or visual and tactile information can be associated with changes in phase coherence (Doesburg et al. 2008; Hipp et al. 2011; Wang et al. 2019). Furthermore, studies in humans have provided evidence for changes of crossmodal coupling during audio-visual speech perception (Lange et al. 2013) and during evaluation of the duration of audiovisual stimuli (van Driel et al. 2014). Evidence is available for the modulation of dynamic crossmodal coupling by factors such as attention, memory, emotion or expectation (Senkowski et al. 2008; Talsma et al. 2010; Senkowski et al. 2014). In the present study, we have addressed the modulation of crossmodal processing in the context of predictions on upcoming sensory stimuli.

A role of neural oscillatory activity in predictive processing has been postulated early on and is supported by work on unimodal processing (for review, see e.g. Engel et al. 2001; Engel and Fries 2010; Arnal and Giraud 2012). It has been suggested that predictions regarding stimulus identity or stimulus timing may involve oscillatory mechanisms in different frequency ranges. While the latter may be associated with low frequency (delta, theta) oscillations, the former might be mediated by interactions predominantly in higher frequency (beta, gamma) bands (Engel and Fries 2010; Arnal and Giraud 2012). The relation between oscillatory activity and predictions has been addressed by only a few studies in the context of crossmodal processing (for review, see Arnal and Giraud 2012; Jessen and Kotz, 2013). While some of these studies

have addressed temporal predictions (Arnal et al. 2015; van Wassenhove and Grzeczkowski 2015), the fast dynamics underlying crossmodal prediction of stimulus sequences has not yet been studied. Furthermore, large-scale dynamic coupling of oscillatory activity across brain areas has not yet been investigated in the context of crossmodal predictions.

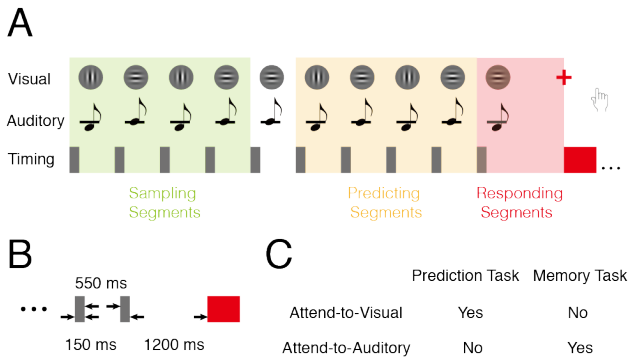

Figure 1: Experimental paradigm. (A) Trial structure: visual gratings (upper row) and auditory tones (middle row) were presented simultaneously in sequential manner (timing in bottom row). The first five items constituted the sample sequence, which would be then repeated in the trial, except for the last item. When the fixation changed its color to red, the subjects had respond by a button press. The data were segmented according to the onset of each stimulus. There were three categories: sampling segments (1st to 4th), predicting segments (6th to 2nd last) and responding segments (last), respectively. (B) Stimulus timing. Each stimulus lasted 150 ms, followed by 550 ms interval; 1200 ms after the last stimulus, the fixation cross turned red. (C) Correct answers for the example shown in (A). In the prediction task, the subjects had to judge whether the last item preserved or violated the sequence in the attended modality; in the memory task, the subjects had to indicate whether the last item was identical to the item before (N-back) in the attended modality. In each block, the subjects performed one of the two tasks and attended to either visual or auditory modality.

Here, we used magnetoencephalography (MEG) to investigate the neural mechanisms underlying prediction of multisensory stimulus sequences. Participants were presented with 5-item-long sequences of audio-visual stimulus pairs which were repeated. At the end of each trial they had to judge whether the last item matched or violated the sequence. Sequence items were always audio-visual, but only one sensory modality was task-relevant on any given trial. As a control, we used a memory task in which participants were requested to judge whether the last item was identical to the item before. We used this experimental design to investigate how intrinsically generated oscillatory signals are involved in predictions of stimulus sequences, and to what extent the coupling between different brain areas may relate to predictive processing. Furthermore, the task design allowed us to test attentional influences on the prediction of multisensory stimulus sequences.

Materials and Methods

Participants

29 healthy volunteers (17 females, age 26.3 ± 4.2 years) participated in this study. They gave written informed consent prior to commencing the experiment. All participants received 10 €/hour remuneration, and additional 30 € bonus if they completed all sessions. All volunteers had normal or corrected-to-normal vision, right handed, and free of psychiatric or neurological disorders. The experiment was carried out in accordance with the guidelines and regulations, approved by the ethics committee of the Medical Association Hamburg.

Experimental procedure

A green fixation cross (0.25˚ visual angle) was shown at the center of the screen, and participants were asked to maintain fixation during the stimulation. A set of 5-item-long binary visual-auditory sequences (32 possibilities) was presented to the participants. Each item appeared for 150 ms, followed by a 550 ms break before the next item. An item was a vertical or horizontal Gabor patch (10˚ visual angle, 0.5 cycles per degree), accompanied by a simultaneously presented sine wave tone which was either high (2000 Hz) or low (1800 Hz) in pitch. The volume of the tone was adjusted to 30 dB above the individual hearing threshold. In each trial, only one of the sequences would be presented and repeated; all sequences were presented with equal probability for all participants. The length of the trials was randomized between 8 and 20 items, with an approximately fixed hazard rate. In the prediction task, participant had to monitor all items and to indicate, by button press, whether the last item of the trial violated the sequence. For different block of trials, participants had to attend to either the auditory or the visual stimuli while ignoring the respective other modality. Although only the stimulus changes in the attended modality defined the task-relevant sequence, the combination of visual and auditory features was kept constant except for the last item (Figure 1). In the

control memory task, the physical stimuli were exactly the same as in the prediction task, but participants had to indicate whether the last item was same as the item second (1-back) or third (2-back) from last in either the visual or the auditory dimension. In the 1-back condition, the participants performed more accurately in the memory task and there was no difference in response time, as compared to the prediction task. In the 2-back condition, the accuracy was the same while the participants were slower in the memory task, as compared to the prediction task. Since there were no substantial difference in the MEG data for these two sets of data, we merged them in the formal analysis. Overall, the study had a two by two design: task (prediction or P, vs. memory, or M) x relevant dimension (visual or V, vs. auditory or A). There were 32 test trials in a block, and each participant would accomplish 16 blocks for one recording session. Each block would be initiated by the participants, who usually took 0.5~2 minutes break between blocks. A mandatory break of about 20 minutes was taken in the middle of each session. The order of blocks in a session were counterbalanced with respect to task and dimension across participants. For each participant two sessions were recorded on separate days, no more than 6 days apart. Before the first MEG recording session, a practice session comprising 4 blocks was conducted outside of the MEG chamber to familiarize participants with the task. The combination of visual and auditory features was different between the practice session and the MEG recordings. Only the data from the MEG sessions were used for subsequent analysis.

The main experiment was conducted in a dimly lit, sound attenuated MEG chamber. The participants were seated in a recording chair with armrest. Visual and auditory stimuli were generated from the same computer (Dell Precision T5500) equipped with Psychtoolbox-3(Brainard DH 1997) and Matlab (R2012b, MathWorks, Natick, MA), on a Windows 7 Professional 64-bit operating system. The visual stimuli were delivered from a projector outside of the chamber via a mirror system to a thin film screen mounted 60 cm in front of the participants' eyes, with 100 Hz refresh rate. The auditory stimuli were delivered with MEG-compatible in-ear headphones (SRM-2525, STAX Limited, Fujimi, Japan). The button presses

were recorded and merged to MEG data with an fORP system (HHSC-2X2, Current Designs Inc., Philadelphia, USA).

MEG recordings and pre-processing

MEG signal was recorded with a 275-channel whole-head system (CTF MEG International Services LP, Coquitlam, Canada) in a magnetically shielded chamber, together with electrical eye, muscle and cardiac activity via Ag/AgCl-electrodes in order to control for endogenous artefacts. Three magnetic sensors fixed at nasion, left and right preauricular points were used to track the head position for both online monitoring (Stolk et al. 2013), artifact rejection and offline data processing. At the beginning of each block, participants would be asked to reposition their head to the original position if the deviation was more than 10 mm. The recorded signals were digitized at a sampling rate of 1200 Hz. After removing malfunctioning channels, we obtained data from 274 sensors for each participant.

Raw data were read into Matlab with functions provided by FieldTrip (Oostenveld et al. 2011). Extreme noisy events (e.g., squid jumps) were detected and data within 1 second epochs around the event were excluded from further analysis. Data were then bandpass filtered between 0.5 ~ 140 Hz with a Butterworth filter at the order of 4, in both forward and reverse direction for zero phase distortion. Both semi-automatic z-score based artifact detection (threshold $z > 7$) and visual inspection were applied to label noisy segments for exclusion. The average head position of each session of each participant was computed and used for later source reconstruction. Epochs where the head position deviated from this position more than 5 mm also labelled excluded. Instead of a notch filter, we removed 50 Hz line noise with sinusoidal estimation (Mitra and Pesaran 1999). The remaining data were submitted to ICA (Hyvarinen 1999) for artifact detection. To decide on the number of ICA components, a PCA was performed. We took the minimum number n, such that the first n PCA components preserved at least 99% of the whole information. Eye movements, cardiac artifacts, remaining 50 Hz line noise and

muscle activities were identified with their distinct spatial extension, spectral signature, waveform properties and distribution across trials (Hipp and Siegel 2013) and then removed from the data.

Event extraction and MEG data segmentation

Time points of stimulus onsets and responses were extracted from the MEG data, and consistency with the events recorded by the stimulus software was verified. Behavioral performance such as hit ratio and response time was calculated based on these events.

The MEG data were segmented based on these events in several categories: (i) *Onset* segments, ranging from -700 to +700 ms relative to the beginning of each trial. This category was used to test neural responses to the physical stimulus properties compared to pre-trial baseline. (ii) *Sampling* segments, ranging from 0 to 700 ms with reference to the onset of the 1st to 4th items of each trial. These segments were used to analyze neural data during acquisition of the sequence. (iii) *Predicting* segments, ranging from 0 to 700 ms relative to the onset of the 6th to the second last item of each trial. This category was used to investigate how the participants predicted the sequences based on the acquired information. We did not include data segments around the 5th item in the analyses because there might be confounding components for both learning and predicting: after the onset of the 5th item, sampling information on this item and predicting the next one could not be disentangled. (iv) *Responding* segments, ranging from 0 to 1500 ms relative to the onset of the last item. These segments were used to analyze response-related aspects. A segment was rejected if it overlapped with any previously labelled data exclusion period. Data were down-sampled to 400Hz after segmenting the cleaning data.

MRI data acquisition and pre-processing

T1-weighted structural MRI scans were obtained for 27 of the 29 participants. MRI recording was performed with a Magnetization Prepared Rapid Gradient Echo (MP RAGE) sequence on

a Siemens 3 T scanner (Magnetom Trio, Siemens, Erlangen, Germany). The acquired images were converted from *Digital Imaging and Communications in Medicine* (DICOM) to *Neuroimaging Informatics Technology Initiative*（NIfTI）format and then processed using the *FreeSurfer* software package (Dale et al. 1999) for cortical segmentation, brain surface inflation and head shape reconstruction. Subsequently, the nasion and the left and right preauricular points were manually labelled on the reconstructed head shape and co-registered with coordinates of these three landmarks obtained from the MEG system, with the aid of the MNE software package (Gramfort et al. 2014). This procedure put the MEG sensors and the structural brain into the same coordinate system.

Source reconstruction

We constructed our source space based on the *FreeSurfer* averaged template brain (Dale et al. 1999). To generate the source space, we started from an icosahedron and then recursively subdivided its edges, the middle point of which would be repositioned to its circumscribed sphere as new vertices. The defined sphere, with 162 vertices on it, was morphed onto the outer surface of the white matter (i.e., the inner surface of the gray matter) of each hemisphere, which yielded 324 locations homogeneously covering the entire cortex. By excluding locations which were unlikely to elicit significant brain responses (e.g., the corpus callosum), we reduced the number of source locations to 290 (Figure S1). This procedure was accomplished with the MNE software package (Gramfort et al. 2014). We chose a relatively low number of source locations for reasons of computational efficiency, and in order to be compatible with the physical spatial resolution of the MEG sensors. These source locations were then mapped to each individual anatomical MRIs. For those participants without an anatomical MRI, we use the template brain ICBM152 from the Montreal Neurological Institute (Mazziotta et al. 1995) instead. As a next step, we computed leadfields with these source locations and individual sensor positions for each session. Leadfield calculation was realized with a single-shell head model (Nolte 2003).

We did not apply compensation for ICA component reduction, with the assumption that the leadfield distribution was not altered by removing components related to eye movements or other muscle activity (Hipp and Siegel 2015). Making use of linearly constrained minimum variance (LCMV) methods (Van Veen et al. 1997), we projected the original sensor-level waveforms to the source locations defined above. For each source location, we computed three orthogonal filters (one for each spatial dimension) that passed activity from the location of interest with unit gain while maximally suppressing activity from all other locations. Next, we linearly combined the three filters into a single filter in the direction of maximal variance (Hipp et al. 2011). In order to avoid spurious effects resulting from unequal filters in between-condition comparisons, we used data from all conditions (with the same number of trials for each condition) to compute a common filter. To derive source estimates, we multiplied the cleaned sensor data with the real-valued filter. Then we merged data from different sessions of the same participant. High source correlations can reduce source amplitudes estimated with beamforming due to source cancellation (Van Veen et al. 1997; Hipp et al. 2011) which might influence the magnitude of cortico-cortical coupling estimates. However, within the range of physiological source correlations (Leopold et al. 2003), the identification of cortico-cortical coupling using the beamforming method should be possible (Kujala et al. 2008). Moreover, although source cancellation might reduce the sensitivity to detect coupling, it should not lead to false positive results.

Statistics

In the analyses of behavioral and MEG data, besides providing significance measures (e.g., *p* values), we also provided measures of effect sizes (η^2_p for ANOVA or cohen's *d* for pairwise comparison).

ERF computation

For each task (prediction, memory) and dimension (visual, auditory) combination (i.e., four conditions), we averaged the waveforms in source space, with trial number stratified for the conditions, to obtain event related fields (ERF) during each type of segment (onset, sampling, predicting, responding).

MEG time-frequency decomposition

The time-frequency decomposition was performed in cortical source space. To avoid computing the same effect twice, we removed the event related fields (ERF) from each data segment prior to the spectral analysis of induced activity. We also computed an averaged ERF of all conditions in the onset segments, from which we evaluated the evoked activity by physical stimulus. All spectral estimates were obtained with the multitaper method (Mitra and Pesaran 1999) and computed across 19 logarithmically scaled frequencies from 5.66 to 128 Hz with 0.25 octave steps. In the induced data, we took time points in steps of 100 ms (the time range of each segment was defined in the *Event extraction and MEG data segmentation*). In the evoked data, we took time points in steps of 25 ms. We characterized responses relative to the pre-stimulus baseline using the bin centered at – 100 ms. The temporal and spectral smoothing was performed as follows. For frequencies larger or equal to 16 Hz, we used temporal windows of 250 ms and adjusted the number of slepian tapers to approximate a spectral smoothing of 3/4 octave; for frequencies lower than 16 Hz, we adjusted the time window to yield a frequency smoothing of 3/4 octaves with a single taper (Hipp et al. 2011). In case the window extended outside of the segment range, zero padding was applied. The employed time-frequency transformation ensured a homogenous sampling and smoothing in time and frequency, as required for subsequent network identification within this multi-dimensional space.

Power estimation and phase coupling analysis

Power of oscillatory signals $X(f, t)$ was estimated as follows:

$$Power_x(f,t) = \frac{1}{n} \cdot \sum_i X_i(f,t) \cdot X_i^*(f,t) \quad (1)$$

Here i is the trial index, t is the time bin index, and f is the frequency bin index, * represents the complex conjugate.

To quantify frequency-dependent synchronization in source space, we estimated the imaginary part of coherency (IMC) – a measure with minimum influence of volume conduction (Nolte G et al. 2004). The IMC between pairs of signals $X(f, t)$ and $Y(f, t)$ was computed according to the following equation:

$$IMC_{xy}(f,t) = Im\left(\frac{1}{n} \cdot \sum_i \frac{X_i(f,t) \cdot Y_i^*(f,t)}{\sqrt{(X_i(f,t) \cdot X_i^*(f,t)) \cdot (Y_i(f,t) \cdot Y_i^*(f,t))}}\right) \quad (2)$$

Here i is the trial index, t is the time bin index, and f is the frequency bin index, * represents the complex conjugate, *Im* indicates imaginary part of complex numbers.

Identification of networks by clustering

The general approach of our network identification closely resembles the method described in (Hipp et al. 2011; Wang et al. 2019) and is summarized below. An interaction between two cortical areas can be represented as a point in two-dimensional space and can be extended to four-dimensional space by adding the dimensions time and frequency. Identifying networks of significant interaction is then equivalent to identifying continuous clusters within this 4-dimensional space. In our approach, we first computed IMCs for all pairs of locations in all time and frequency bins and conditions with equation (2). Then t-statistics (for pair-wise comparison) or ANOVA (for multiple-factor analysis) was performed across participants. The resulting multi-dimensional matrix was then thresholded with a t-value or F-value (e.g. equivalent to $p = 0.01$), setting every connection to either 1 (connected) or 0 (not connected). Further, we performed a neighborhood filtering using threshold (e.g. 0.3) to remove spurious

connections: if the existing neighboring connections of a certain connection were fewer than (e.g. 30%) of all its possible neighboring connections, this connection would be removed. As one connection was defined by four numbers, representing its frequency, time, and the two locations of a pair, two connections were defined as neighbors if they were identical in three of the four numbers, and the other value would only differ by one step, i.e., 100 ms in time, 0.25 octaves in frequency, or immediate neighborhood in source space. Two source locations were defined as immediate neighbors if they were directly connected in the regular polyhedron during source space construction. Eventually, the remaining neighbor connections formed the candidate networks for statistics. We defined their sizes as the integral of the *t*-values or *F*-values across the volume, time and frequency. To evaluate their significance, random permutation statistics was performed to account for multiple comparisons across the interaction space. To this end, the network-identification approach was repeated 1000 times with shuffled condition labels to create an empirical distribution of network sizes under the null-hypothesis of no difference between conditions (Nichols and Holmes 2002). The null-distribution was constructed from the largest networks (two-tailed) of each resample, and only networks with sizes ranked top 5% in these null distributions were considered as significant (i.e. $p = 0.05$) in the comparison. The final networks corresponded to cortical regions with different synchronization status among comparisons, and were continuous across time, frequency, and pairwise space. Applying this method, we compensated for the relatively arbitrary thresholding procedure during the generation of the 4-D connectivity array: the networks with fewer statistically stronger connections and networks with more statistically weaker connections were both considered.

The identification of power difference network was similar to the above except that there was only one dimension in space rather than two for coupling.

Identification of clusters for various contrasts

With the methods described above, we applied the cluster identification approach for analysis of several contrasts. We identified the cluster of regions showing stimulus-related power changes by contrasting the evoked power from each time bin between 0 and 700 ms after the onset of the first item of each trial, with the power averaged between time bins between 125 and 75 ms before stimulus onset, with a pairwise *t*-test. Here, we use a threshold of *t* values corresponding to $p = 0.01$; the threshold for the neighborhood spatial filter was 0.3, i.e., at least 30% of neighboring difference had to be significant.

To investigate differences of induced oscillation power between conditions, we employed a 2 x 2 design ANOVA, with factors task (prediction vs. memory) and attended modality (visual vs. auditory). Here we use a threshold of *F* values corresponding to $p = 0.05$ for the main effects and interactions; the threshold for the neighborhood spatial filter was 0.3, i.e., a test with significant difference would be regard as significant after filtering, when at least 30% of the tests, which included data differ only one bin in time or frequency, or only one location was different but still was immediate neighbor of the data in the inspected test, were also significant. i.e. close to orphan significant points would be removed. to A post-hoc test was applied for each component of the network to test the direction of difference. If directions of effects were not consistent within a cluster, the cluster was split to separate the positive and negative effects. The same approach was used to investigate differences of coupling between conditions.

In order to investigate the effects of stimulus context, we classified the segments according to its previous or next stimulus, which could be the same or different. With this factor and the attended dimension, we construct another 2 x 2 ANOVA design. This was submitted to the network identification with the threshold of *F* values corresponding to $p = 0.01$ for the main effects and interactions; the threshold for the neighborhood spatial filter was set to 0.3, i.e., at least 30% of neighboring difference had to be significant. Again, a post-hoc test was applied

for each component of the network to test the direction of the difference. If appropriate, the network was split to separate the positive and negative effects.

Furthermore, we tested both the power and coupling difference between sequence preservation and violation using paired *t*-test and the results were submitted to the clustering approach with the threshold of *t* values corresponding to $p = 0.01$; in this case, the threshold of neighborhood spatial filter was also set to 0.3.

Illustration of identified clusters

To visualize networks identified by the clustering approach, we projected them separately onto two subspaces. We integrated the *t*-values or *F*-values among connections over all spatial locations for each time and frequency bin to illustrate when and at which frequencies a network was active irrespective of the spatial distribution. Complementary to this, we also computed the integral of *t*-values in the connection space over time, frequency, and target locations (only for coupling networks) for each location in the corresponding network. The result was then displayed on the template brain surface to reveal the spatial extent of the network independent of its intrinsic coupling structure and time-frequency characteristics. For illustration purposes, the data were interpolated from 290 to ~300,000 vertices on the inflated brain surface. The labeling of brain areas involved in our networks was based on their relative position to the gyri and sulci of the template brain (Fischl et al., 2004).

Performance modulation by alpha phase within the attention network

In order to test whether the phase of alpha oscillations would alternate performance we first assigned the response times of each trial to all of its segments. We then split the phases evenly into bins with a width of $\pi/8$. For each component (as specified by a time bin, a frequency bin and a location or location pair) of the attention network observed during the sampling segments (Figure 4A), we assigned the segments by their phase within this component to the phase bins;

their corresponding response times were then averaged within the component and subsequently over the whole network. Thus, we obtained an 8-point vector for each condition and participant. The data were normalized by subtracting the minimum and then dividing by the difference between maximum and minimum. The normalized data were fit with a cosine function $y = (1 - \cos(x - \theta))/2$, where θ was the optimal phase at which we expected to obtain the minimum response time. We fitted the parameter θ for all conditions (attend-to-visual and attend-to-auditory) and participants, and tested the difference between conditions with a paired *t*-test. Note that due to the periodicity (2π) of the cosine function, the difference exceeding the range of $(-\pi, \pi)$ would be modified to fit into this range: e.g., 1.7π would be converted to -0.3π.

To investigate differences of the phase effects between conditions, we compared the range of modulation (difference between maximum and minimum) for the two conditions using a paired *t*-test.

To compare effects between clusters obtained predicting segments, and then calculate the integral across the whole cluster to yield one number for every participant, which was then submitted to a paired *t*-test with the original values (i.e. same procedure but with data in sampling segments) of this cluster. This test indicate that if this cluster were in predicting segments, would the total alpha power be larger or smaller. Similarly, we did the test for the with the same contrast in different sets of data (sampling segments vs. predicting segments, e.g., cluster in Figure 4A and B), we took each component, as identified in a certain time and frequency bin and location, of cluster yielded from the sampling segments, and supply data from other cluster (generated from data in predicting segments), and the components shared by the two clusters.

Performance modulation by theta phase within the prediction network

We tested the modulation of behavioral performance by the phase of the coupling in the theta band networks related to prediction (Figure 6 C, D). We defined the optimal phase of the

coupling by averaging the segments used to quantify the network (all correct and stratified for conditions) for each component. Each segment was classified into one of the eight bins based its delay to the optimal phase. We averaged the response time from the same data for each component and then averaged these across the whole network. Since here the delay was a relative measure (difference to optimal), we averaged the tuning of response time by phase across participants.

We performed a similar analysis as the above to test for effects on accuracy, with two main differences as follows. First, we had to include all the data for computation the percentage of hits, not only the data that were used to generate the network (correct and stratified). Second, we used half of the correct data to generate the optimal phase and classify half of the full data, which would not include the trials to generate the optimal phase, and this was done as well for the respective other half of the data.

To rule out an influence of the number of segments, which was significantly larger for zero phase delay than a delay of π, we re-classified the data by absolute values of phase delay to the optimal phase, where the spacing was not exactly evenly distributed, e.g. the data range was smallest for the bin closest to zero so that it could include same number of segments as the bin farthest from zero, which was much wider. Then a non-parametric correlation was computed to test whether the data followed the same trends. Spearman's rho was calculated.

We also took the top and bottom 25% of data based on the absolute values of difference to optimal phase, and then compared the difference of performance with a paired t-test. We pooled the *t*-values for each voxel. Then we also took the *F*-values of each voxel when generating the network. The two group of values were correlated, and Pearson's *r* was measured.

The distance of coupling and its dependence on oscillation frequency

In the networks obtained by contrasting coupling differences, we tested the distance between the coupled locations. Here we defined the distance as the topological distance, i.e., if two

locations were immediate neighbors in the source space constructing process (see section *Source reconstruction*), the distance was defined as 1, otherwise if they were not but they shared one voxel as immediate neighbors, the distance was defined as 2, and so forth.

Then we computed the average distances of all pairs of connections in the coupling networks (Figure 6 C-F), and applied a 2 x 2 ANOVA with factors network dominance (prediction > memory or memory > prediction) and data set (sampling segments vs. predicting segments).

We also pooled the connections of all networks and classified two groups, oscillating in the gamma band (frequency range 53.8 – 128 Hz) or theta/alpha band (frequency range 5.7 – 11.3 Hz); the coupling distance distributions were examined for these two groups.

Cluster analysis based on stimulus context

As mentioned in section *Identification of clusters for various contrasts*, we detected two clusters related to stimulus context, i.e., occurrence of the same versus different stimulus as the previous or next item respectively. We calculated modulation index (MI) by dividing the difference of same and different condition and then dividing by their sum. MIs were also computed for the corresponding data for sampling segments and memory task, and compared using a 2 x 2 ANOVA.

Interactions among clusters and cross-frequency coupling

To investigate interactions among clusters, especially between the theta coupling cluster and the clusters defined by stimulus context, we computed cross-frequency coupling. We first chose the locations that were most significant in both clusters, i.e., at least 1 stand deviation higher than the mean for the integrated *F*-values. Then we computed the theta (4 - 7 Hz) phase of these voxels in the theta networks and the beta (12 - 32 Hz) power of these voxels in the beta clusters. The cross-frequency coupling among these times courses was computed with the modulation index (Tort et al. 2010; Wang et al. 2019). As a control, we computed a shift predictor by

shuffling the trials, such that the phase from trial A would be coupled with trial B, with all other intact. In this case, the stimulus locked information would be preserved but instantaneous correspondence of phase and amplitude was removed. Subsequently, the cross-frequency coupling of real data was normalized with those obtained from shift predictors. Then they were compared across conditions with ANOVA.

Results

Behavioral data

Participants performed the task with an average hit ratio of 90.5%. Both the hit ratio and response time were submitted to a 2 by 2 analysis of variance (ANOVA), with two factors: task (prediction or memory) and attended modality (visual or auditory). For the hit ratio, neither the interaction between the two factors nor the main effects of factor modality were significant. However, the main effect of factor task was significant (F (1, 28) = 6.80, p = 0.01): the hit ratio was higher in memory task (η^2_p = 0.20). There was significant interaction between the factor task and factor dimension in the response time (F (1, 28) = 13.31, p = 0.001, η^2_p = 0.32). Post-hoc analysis showed that, in the prediction task, responses were slower (p = 0.005, cohen's d = 2.21) when participants attended to the auditory modality; in the memory task, the difference was even larger (p < 0.001, cohen's d = 2.27) in the same direction. The main effect of factor task was also significant (F (1, 28) = 4.52, p = 0.04): the responses were faster in the prediction task (η^2_p = 0.14).

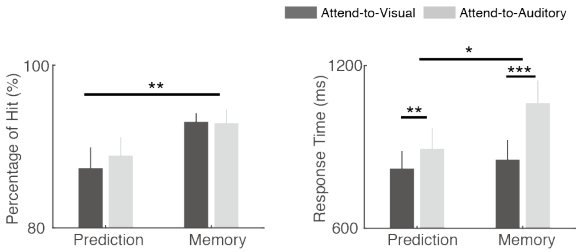

Figure 2: Behavioral performance. Mean and its standard error for both percentage of hit (left) and response time (right) were compared across two tasks and attended modalities. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$

Audiovisual stimuli lead to power changes in multiple frequency ranges

Since in all conditions we always presented similar audio-visual stimuli (gratings accompanied by tones), we pooled the data across conditions with stratified trial numbers and examined the power change caused by the first stimuli of the sequences, as compared to baseline (-100 ms before trial onset) with the clustering approach. The results are illustrated in Figure 3. The stimuli generally caused a power increase in most of the brain areas, which was most salient in

bilateral occipital areas, fusiform gyri, transverse temporal gyri, superior temporal gyri, superior temporal sulci, intraparietal sulci, precunei and cingulate cortices. The stimulus-related activation was broadband in frequency, with gamma power changes being larger than expected from a power law distribution. The activations were strongest after stimulus onset, and showed a second peak after stimulus offset.

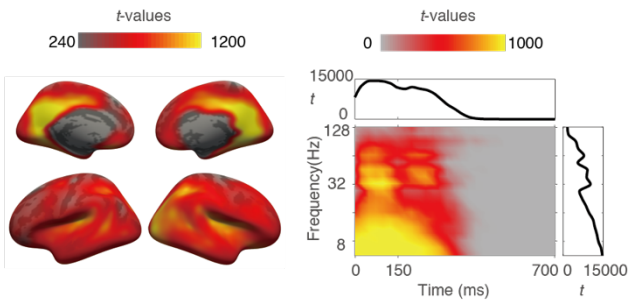

Figure 3: Stimulus-related power changes. (Left) Spatial distribution of the cluster. Color is scaled to the integral of *t* values across time bins and frequency bins for each voxel. (Right) Time-frequency distribution of the network; color is scaled to the integral of *t* values across all voxels for each time and frequency bin. Inset plots show integrated *t* values over time (top, x-axis: time in ms, y-axis: *t* value) or frequency (right, x-axis: *t* value, y-axis: frequency in Hz).

Attention to auditory stimuli induced more alpha power than attention to the visual modality

In addition to the stimulus-related power changes relative to the pre-trial baseline, which reflected the general processing of the audio-visual stimuli shared by all conditions, we also examined power differences between conditions. To investigate the effects of crossmodal attention and prediction, a 2 x 2 ANOVA design with factors task (prediction vs. memory) and attended modalities (visual vs. auditory) was employed. The obtained multi-dimensional array of *F*-values was submitted to a clustering analysis (see method for details). This analysis revealed clusters of source locations showing main effects for both factors, but no significant interactions (Figure 4). The attention effects were observed both in the sampling segments (permutation test, $p = 0.002$, Figure 4A) and in the predicting segments (permutation test, $p < 0.001$, Figure 4B), which were quite similar to each other in both spatial and spectro-temporal distribution. Significant differences were observed in the alpha and low beta range across almost the entire segment and peaking at about 400 ms after stimulus onset. The attention effect

mainly localized to lateral occipital cortices, superior parietal cortices, precunei and insulae of both hemispheres. Post-hoc tests revealed that alpha power was always stronger when participants attended to the auditory modality. As alpha oscillations are thought to be related to functional inhibition (Jensen and Mazaheri 2010), this suggests that the processing of visual information was inhibited in trials where auditory information was relevant. However, we did not observe the opposite effect, i.e., stronger alpha power in auditory areas during attention to the visual modality.

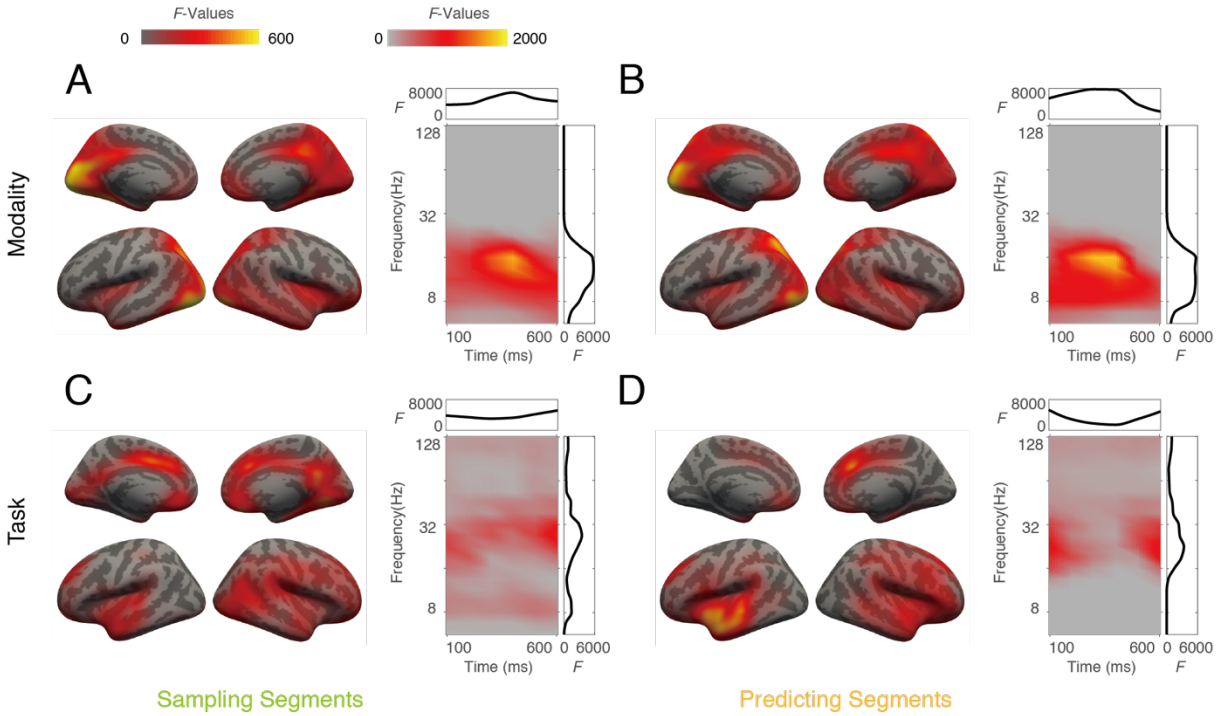

Figure 4: Power differences across conditions. The clusters were obtained with 2 x 2 ANOVA design on oscillation power, during both sampling (A, C) and predicting (B, D) segments. Significant main effects of factor modality (A, B) and task (C, D) were observed. For each cluster, both spatial (left) and time-frequency (right) distributions are illustrated. Color indicates accumulated *F* values. Side plots in box are integrated *F* values plotted against time (top, x-axis: time in ms, y-axis: *F* value) or frequency (right, x-axis: *F* value, y-axis: frequency in Hz).

Prediction task induced less beta power than memory task

We also observed main effects of power differences between prediction and memory tasks in both sampling (permutation test, $p = 0.006$, Figure 4C) and predicting segments (permutation test, $p = 0.02$, Figure 4D). These clusters involved bilateral superior temporal gyri, fusiform

gyri, precentral gyri, insula, cingulate and frontal cortices. Further analysis revealed that power in this cluster was always stronger in the memory task. In the sampling segments, significant differences were observed in the alpha and beta band. In the predicting segments, the alpha component was absent. This alpha component was mainly observed in occipital areas.

Response time was modulated by alpha phase in sampling segments

It has been suggested that alpha oscillations may function as gain controller, i.e., processing of information arriving at certain phases of the alpha oscillation will be facilitated and otherwise inhibited (Jensen and Mazaheri 2010). If this holds true, processing efficiency should depend on the phase of the alpha oscillation, and this phase relation should differ between the attend-to-visual and the attend-to-auditory condition. We hypothesized that when attention was directed to auditory information, there should be an optimal phase of the alpha oscillation where the performance should be better than at other phases; at the same phase, the processing of visual information should be less efficient; and vice versa. We tested this hypothesis as follows. Based on the results of the first cluster (Figure 4A), we sorted the normalized response times associated to each data segment according to its oscillation phase, and then pooled the data for the attend-to-visual and attend-to-auditory conditions separately. We observed that the response times were modulated differently in the two conditions (Figure 5A, B). By fitting the modulation with a cosine function, we estimated the phase related to best performance in both conditions for each participant. These phases were significantly different ($p = 0.005$, cohen's $d = 0.43$) between the attend-to-visual and attend-to-auditory conditions (Figure 5 C), suggesting a functional significance of the alpha oscillation phase in this cluster. Furthermore, the modulation range of attend-to-auditory should be larger, since in this case the alpha oscillations were stronger. This was confirmed by the data ($p = 0.006$, cohen's $d = 0.41$, Figure 5 D). We did not observe similar effects in the cluster shown in Figure 4B, which might indicate that in

the predicting segments, the alpha oscillations may not be as strong as in the sampling segments. Thus, we estimated the integral alpha power in predicting segments for the components of the cluster obtained during the sampling segments (Figure 4A), and compared the result with its counterpart in sampling segments; we also analyzed the reverse and, moreover, we also quantified the integral of alpha power in both groups of segments within components that are shared by the two clusters. The results support our hypothesis that the power was always smaller in predicting segments for these comparisons (p = 0.018, cohen's d = 0.46 for cluster from sampling segments; p = 0.005, cohen's d = 0.57 for cluster from predicting segments; p = 0.018, cohen's d = 0.47 for their shared components, Figure 5E).

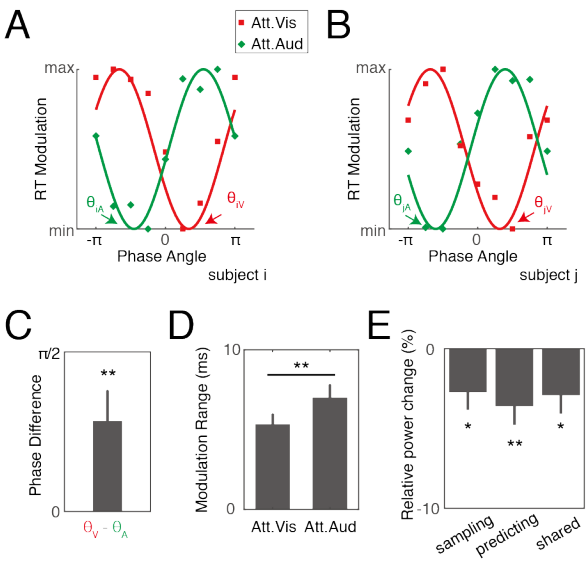

Figure 5: **Dependence of performance on alpha oscillation phase**. (A) and (B) are examples for results from two representative participants. Data segments were classified by the phase difference to the average phase of all analyzed segments. The response times (RTs) corresponding to all segments aligned to a certain phase bin were averaged. The averaged response times were normalized for each condition and participant, and then fit with a cosine function, yielding an estimate of the optimal phase for each condition in a certain participant. (C) Mean difference between optimal phases in attend-to-visual and attend-to-auditory condition. (D) Comparison of modulation range between two attended modalities. (E) The relative power differences between two type of segments. In C-E, error bars represent stand error. *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$.

Prediction task was associated with stronger theta coupling and weaker gamma coupling

As a next step, we analyzed functional connectivity, as measured by the imaginary part of coherency (Nolte et al. 2004), which minimized the influence of volume conduction. In this dataset we employed the same ANOVA design as for the power analysis in the preceding section. We observed significant coupling differences for the factor task (prediction vs. memory) in both sampling (permutation test, p = 0.018, Figure 6A) and predicting segments (permutation test, p = 0.026, Figure 6B). The clusters were similar for both segment types and extended to superior temporal gyri, pre- and postcentral cortices, cingulate cortices, prefrontal areas, insulae

and precunei in both hemispheres. Coupling differences between the two tasks were observed in the theta, alpha and gamma ranges. Interestingly, further analysis revealed that the theta and

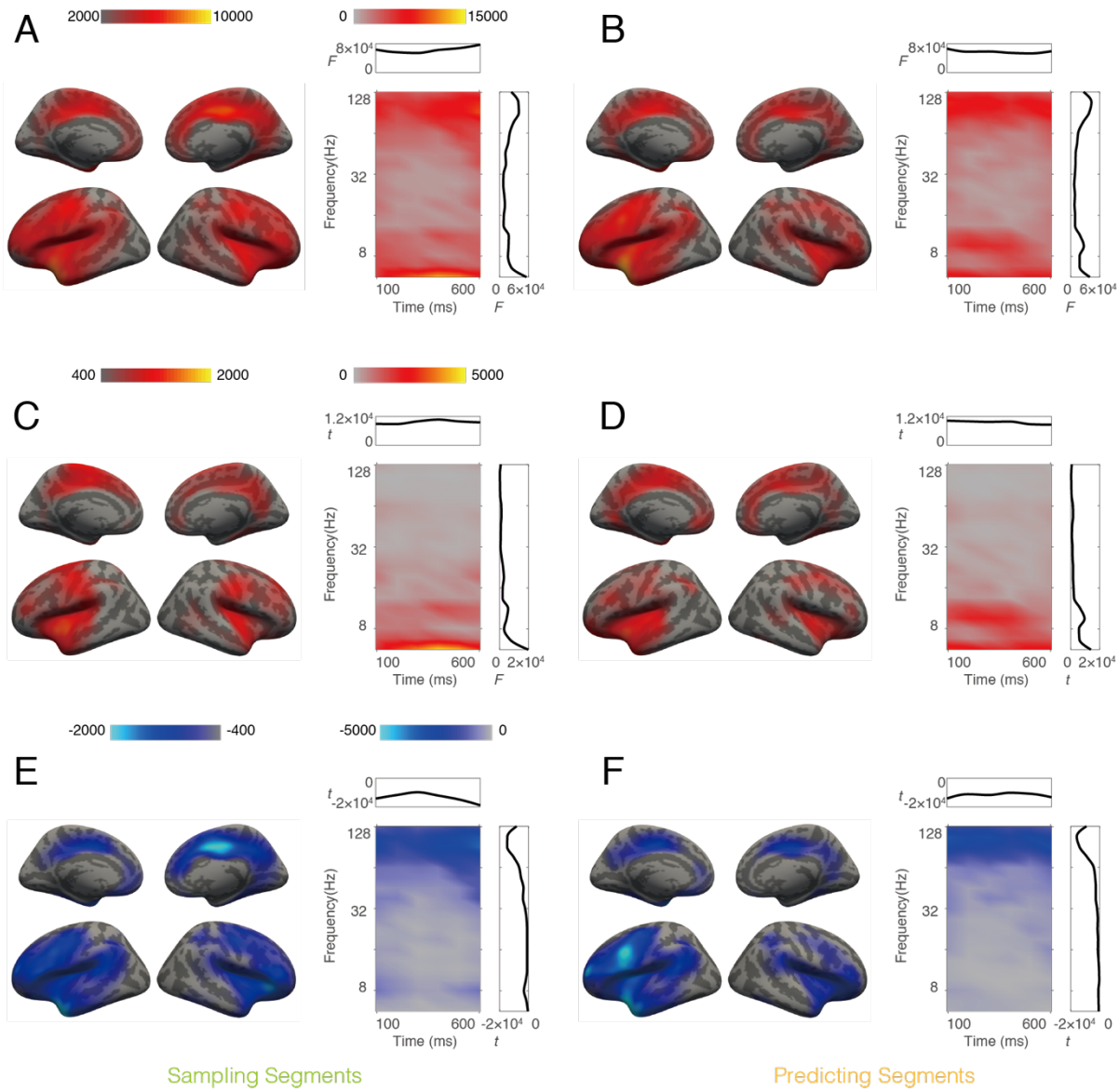

Figure 6: Coupling differences between prediction and memory tasks. (A, B) Main effects of factor task in a 2 x 2 ANOVA design. (C-F) Results of paired *t*-test for the same contrast. Left column shows data from sampling segments and right column shows data from predicting segments. In each panel, the spatial distribution of the cluster is displayed on the left. Color is scaled to integral of *F* or *t* values across time bins, frequency bins and connected locations for each voxel. On the right, the time-frequency distribution of the cluster are shown, color is scaled to integral of *F* or *t* values across voxel pairs for each time and frequency bin. Side plots in box are integrated *t* values against time (top, x-axis: time in ms, y-axis: *F* or *t* value) or frequency (right, x-axis: *F* or *t* value, y-axis: frequency in Hz). Color bars are shared between A, B, between C, D, and between E, F.

alpha coupling was stronger in the prediction task (Figure 6 C, D), whereas gamma coupling was stronger for the memory task (Figure 6 E, F). The spatial distribution of low frequency components was similar to that of the high frequency components. It seemed that in these brain

areas, theta and alpha coupling favorites the prediction task and gamma coupling favorites the memory task.

Cross regional theta phase delay modulated both accuracy and response time

It has been suggested that brain areas communicate by coherent oscillations if the interaction occurs at an optimal phase delay (Womelsdorf et al. 2007; Fries 2015). Evidence supporting this hypothesis has been obtained in neurophysiological recordings in animal studies at the level of single neurons or small neuronal populations. In the present dataset, we tested whether this notion is also supported by analysis of large scale data recorded with MEG from the human brain. In the theta clusters we obtained (Figure 6 C, D), we determined the optimal phase for each component (as defined by a certain location pair, time and frequency bin) as the average phase lag between the location pair at this time and frequency bin, in all segments corresponding to the data used to generate them (correct trials stratified for each condition). We then classified the phase delay to the optimal phase into 8 bins (from $-\pi$ to π), for which the performance data (response time in the same dataset, and percentage of hit with all data included) were pooled and averaged. Thus, we could obtain the modulation of performance by phase delays for each cluster (Figure 7 A-D). The results suggested that in both sampling and predicting segments, the percentage of hit was highest at the zero delay to optimal phase, and then decreased with the delay, in the shape of a cosine function for the prediction task (Figure 7A, C, red). Similarly, the response time of prediction task was shortest in the zero delay and then increased with phase delay for the predicting segments - the modulation resembled an inverted cosine function (Figure 7D, red). The response time modulation in the prediction task during sampling segments showed the same trends but was less significant (Figure 7B, red). The modulation of percentage of hit and response time in the memory task exhibited a similar but much weaker pattern in the predicting segments (Figure 7C, D, green), but not in the sampling segments. In the optimal

phase, there were more data segments for the bin with zeros delay, which were fewer in larger delay (e.g. -π and π). To rule out the possibility that the difference of phase modulation was due

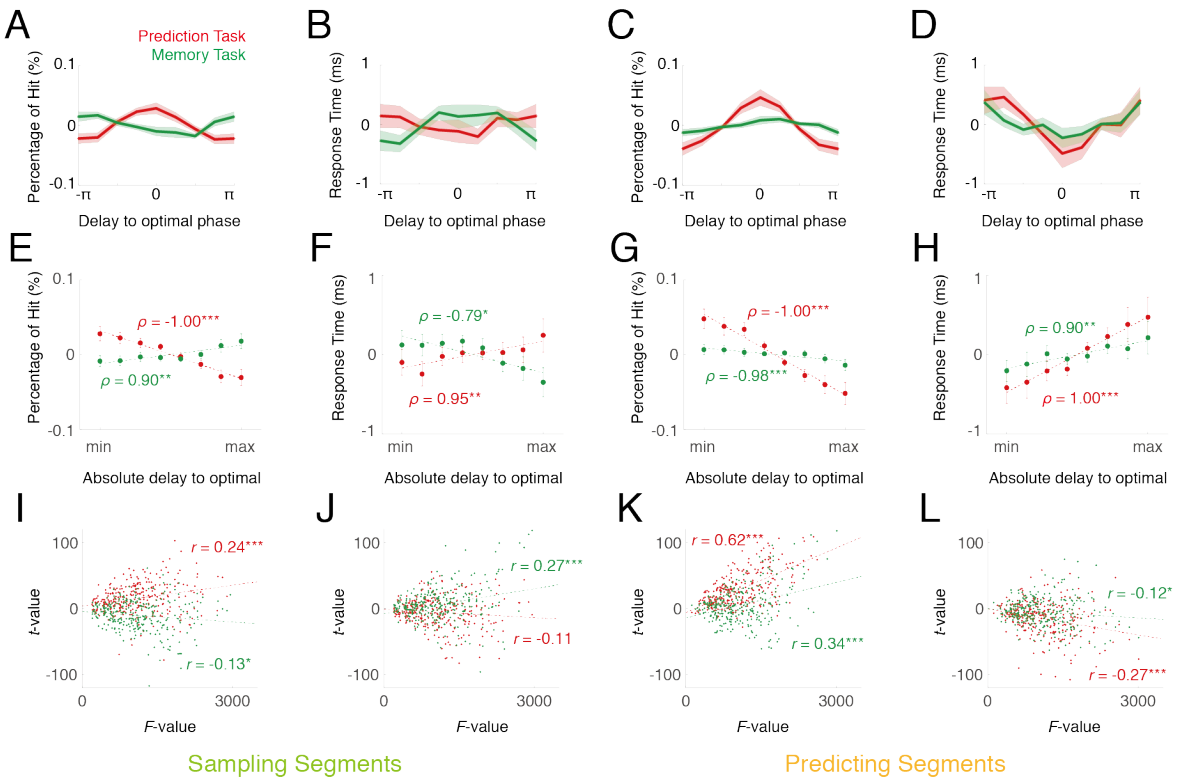

Figure 7: Dependence of behavior on theta coupling phase. Top row: dependence of percentage of hits (A, C) and response time (B, D) on phase delay to the optimal phase. Middle row: percentage of hit (E, G) and response time (F, H) dependence on absolute delay to the optimal phase, with equal number of segments for each data point. Error bar indicates stand error. Spearman's correlation coefficient (ρ) is labeled. In panel A – H the mean was removed for all percentages of hit and response times. Bottom row: Correlation of statistical power for percentage of hit (I, K) and response time (J, L) difference and statistical power of cluster among the voxels. Pearson's correlation coefficient (r) was labelled on the plots. The left two columns show data from the sampling segments and the right two columns show data from the predicting segments. *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$

to amount of data, we sorted the segments into 8 bins, according to absolute delay to the optimal phase, such that within each bin, there were equal number of segments. Subsequently, we submitted the performance per bin to a non-parametric correlation (Figure 7 E-H), the results of which suggested the same, i.e., the effects were not caused by variations in the amount of data.

We further took the performance of data with top 25% delay and bottom 25% delay for each voxel, and performed a paired *t*-test. These results were correlated with the integral *F*-values of each voxel from the clustering approach, which yielded significant positive correlation (Figure

7 I - L). This analysis suggested that, the more likely a brain location was part of the theta cluster, the stronger the behavioral modulation by the theta phase.

Theta cluster involved coupling over larger distances than gamma cluster

Previous studies have suggested that oscillations with lower frequency may mediate interactions over larger distances (Kopell et al. 2000; von Stein and Sarnthein 2000). We thus further examined the distances of the interacting locations within the coupling clusters, and tested the relation to the frequency ranges. The distance of two coupling nodes was defined by smallest number of immediate neighbor connection needed to link them in the topological anatomical structure (see star methods). We found that in the theta clusters, the average distances were larger than those in the gamma clusters (Figure S2A) for both sampling segments ($p < 0.001$, cohen's $d = 0.25$) and predicting segments ($p < 0.001$, cohen's $d = 0.19$). This suggests that functional connections supporting the prediction task may involve interactions over larger spatial distances. We also tested whether the coupling distance was dependent on oscillation frequencies. To this end, we pooled all connection of the clusters and tested the distribution of theta/alpha frequency and gamma frequency (Figure S2B). The results suggested that coupling in lower frequency ranges was associated with larger interaction distances. Specifically, the clusters supporting sequence prediction task involved a larger fraction of long range connections.

Repetition of stimulus items induced stronger beta power

The prediction effects observed in the prediction task might potentially also be influenced by context effects resulting from repetition of stimulus items. To test this possibility, we generated a contrast defined by the stimulus context. Each segment was classified by its relation to the

preceding segment, which could be the same or different, in the respective attended modality. This yielded a 2 x 2 ANOVA with the factors modality (auditory vs. visual) and context (same

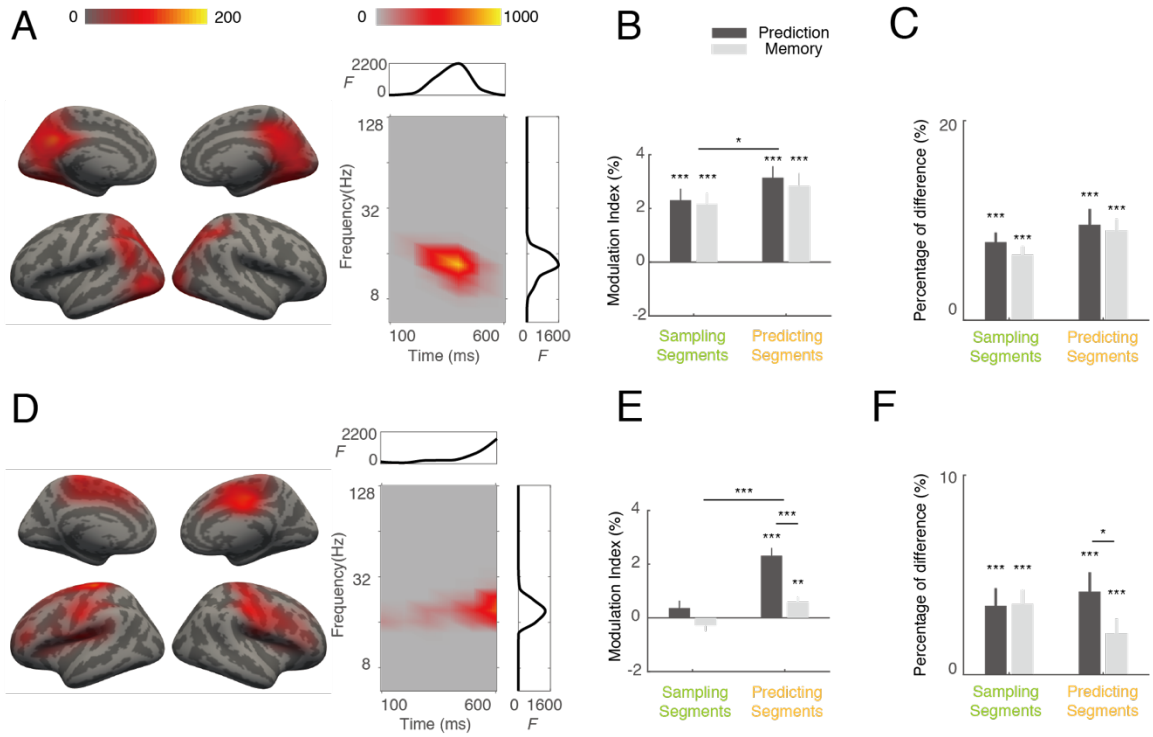

Figure 8: Power differences as a function of context. (A - C) Effects of same vs. different preceding item. (D - F) Effects of same vs. different subsequent item. Left column (A, D): spatial (left) and time-frequency (right) distributions of the cluster are illustrated. Color indicates accumulated *F* values. Side plots in box are integrated *F* values against time (top, x-axis: time in ms, y-axis: *F* value) or frequency (right, x-axis: *F* value, y-axis: frequency in Hz). Color bars are same for panel A and C. Middle column (B, E): Modulation indices. Right column (C, F): Phase amplitude coupling between significant locations shared between coupling cluster and the context clusters, as changes to that of the shift predictor. *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$

vs. different). We applied our clustering approach with this design, and found one cluster for the main effect of context (relation to previous item, permutation test, $p < 0.001$, Figure 8A). The cluster mainly involved lateral occipital areas, intraparietal areas and precunei. It peaked around 400 ms in time and 13.5 Hz in frequency. Further analysis suggested that the power was stronger for the 'same' condition than the 'different' condition. We defined a modulation index (MI) as (same – different) / (same + different). The MIs were always larger than zero (all four comparisons $p < 0.001$, cohen's $d = 1.41$ for prediction task in sampling segments, 1.32 for memory task in sampling segments, 1.95 for prediction task in predicting segments, 1.53 for memory task in predicting segments, Figure 8B), and were stronger during predicting segments

than in sampling segments (F (1, 28) = 4.95, p = 0.034, η^2_p = 0.15, Figure 8B). These results suggested that the information on preceding stimuli could be temporarily stored in this network. With a similar analysis, we obtained another cluster for the main effects of context (relation to the next item (permutation test, p < 0.001, Figure 8D). It mainly activated the pre- and postcentral areas, supplementary motor areas, and lateral prefrontal areas. It peaked around 600 ms after stimulus onset (or 100 ms before the next stimulus), and 19.0 Hz. Further analysis suggested that the power was stronger for the 'same' condition than the 'different' condition. MIs of this cluster were not significantly different from zero in sampling segments, and were significantly larger in predicting segments (p < 0.001, cohen's d = 2.17 for prediction task, p = 0.007, cohen's d = 0.76, Figure 8E). It was stronger for the prediction task than memory task (F (1, 28) = 27.01, p < 0.001, η^2_p = 0.49, Figure 8E). This suggests that, before onset of the next stimulus, activity in this network might encode the forthcoming sequence items.

Multiple clusters are linked by phase amplitude coupling

A question of obvious interest is how these different clusters interact with each other. One possibility is that the clusters reflecting the context effects of preceding items might interact with the theta coupling cluster to form the sequence information. The sequence information could, in turn, influence the cluster reflecting the contextual prediction of the next item. To test this idea, we examined cross frequency coupling for the nodes shared by these clusters, i.e., whether the theta phase of the coupling cluster modulated the beta power of the context cluster. We observed that in all cases the cross-frequency coupling were stronger than a shift predictor computed as a control (Figure 8 C). Only in the predicting segments the cross-frequency coupling was stronger for the prediction task than the memory task (Figure 8 F).

Sequence preservation vs. violation induced differences in both power and coupling

We examined power differences between sequence preservation and violation for the responding segments with paired *t*-test. The results are shown in Figure S3. The first cluster (permutation test, $p < 0.001$, Figure S3A) exhibits stronger power for sequence preservation. The major components were in theta-alpha range and peaked around 900 ms. The activated areas in this cluster include precunei, cingulate cortices, pre-and postcentral areas, frontal areas, superior temporal gyri, middle temporal sulci, transverse temporal gyri, intraparietal sulci and insulae in both hemispheres.

The second cluster (permutation test, $p < 0.001$, Figure S3B) included anterior cingulate cortices in both hemispheres and superior temporal cortex, insular, pre-and postcentral areas, inferior precentral sulcus in the right hemisphere, where the theta power was stronger for sequence violation after stimulus onset, which peaked around 300 ms.

The third cluster (permutation test, $p < 0.001$, Figure S3C) included middle cingulate cortex, superior temporal cortex, insular, pre-and postcentral areas in the right hemisphere, where the beta power was stronger for sequence violation just before the response cue, which peaked around 1100 ms after stimulus onset.

We examined the coupling difference, as measured by imaginary part of coherency, between sequence preservation and violation for the responding segments with a paired *t*-test. The results are shown in Figure S4. The first cluster (permutation test, $p = 0.001$, Figure S4A) exhibits stronger coupling for stimulus preservation in the theta range, peaking around 1000 ms. The activated areas in this cluster included middle cingulate cortices, pre-and postcentral areas and transverse temporal gyri in both hemispheres, and superior temporal gyri, insulae and superior frontal gyri mainly in the right hemisphere.

The second cluster (permutation test, $p = 0.005$, Figure S4B) included middle cingulate cortices, pre- and postcentral areas, inferior frontal areas and insulae in both hemispheres and superior

temporal cortex mainly in the right hemisphere, where the theta coupling was stronger for sequence violation after stimulus onset, which peaked around 400 ms.

Discussion

In the present study, we have recorded MEG during a serial prediction task in which participants had to acquire stimulus sequences and to monitor whether subsequent probe items complied with the sequence. Presentation of the audiovisual stimulus items revealed activation in a widely distributed set of areas involving early and higher-order visual and auditory areas as well as intraparietal sulci, precunei and cingulate cortices. Task-related changes of power and coupling of neural oscillations were observed in several frequency bands. We found prediction-related theta-band coupling in a network involving cingulate, premotor, prefrontal and superior temporal regions. Behavioral performance of the participants correlated with phase delays between remote regions in this network. In this network, beta power was stronger when the previous or next stimuli were the same as the current one in the sequence. In the prediction network, we observed cross-frequency coupling between theta and beta signals. We also observed asymmetrical effects of attention to the visual and auditory modalities in a network distinguished by changes in alpha power, where alpha phases modulated response times. Overall, our results suggest that oscillations in multiple frequency ranges as well as coupling within and across frequencies may have an important role in sequence processing.

Attention-related network changes

In the task employed in this study, features in both modalities varied along each trial, of which only one, either in the visual or auditory modality, was task-relevant while changes in the other modality had to be ignored. We therefore expected attention-related changes when contrasting the two conditions in our data-driven power analysis. Indeed, we did obtain a network related to the intermodal attention shift characterized by alpha-band power changes, which involved the lateral occipital cortices, superior parietal cortices, precunei and insulae of both hemispheres. The pattern of activated areas is in line with earlier studies on crossmodal attention shifts

(Senkowski et al 2008; Friese et al 2016; Misselhorn et al 2019). The insula is part of the saliency network and has been suggested to be involved in attentional processing and in initiation of cognitive control (Menon and Uddin 2010). The precuneus is a hub in the default mode network, which also closely interacts with the frontal-parietal attention network (Utevsky et al. 2014). Further analysis revealed that the power in this network was stronger when participants attended to the auditory modality, suggesting that alpha oscillations were related to functional inhibition of visual areas. We did not observe the counterpart of these effects, i.e., stronger alpha power in auditory regions during attention to the visual modality. This asymmetry might be due to an intrinsically stronger perceptual dominance of vision; thus, no inhibition of its competitor may have been needed in our task setting.

It has been suggested that alpha oscillations may function as gain controller, i.e., processing of information arriving at certain phases of the alpha oscillation will be facilitated (Jensen and Mazaheri 2010; Foxe and Snyder 2011; Jensen et al. 2014; Womelsdorf et al. 2014; VanRullen 2016). We hypothesized that, if this holds true, processing efficiency should depend on the phase of the alpha oscillation, and this phase relation should differ between the attend-to-visual and the attend-to-auditory condition. For the alpha-band cluster we sorted the reaction times according to the oscillation phase of each data point, and then pooled the data for attend-to-visual and attend-to-auditory conditions separately. In line with our hypothesis, we observed that the optimal phases for response time were significantly different between the attend-to-visual and attend-to-auditory conditions, supporting the notion that alpha oscillations have a functional significance for gating of sensory information. These observations are in line with other studies demonstrating a dependence of behavioral performance on the phase of alpha oscillations. In one study, distributions of alpha phase were compared for hits and misses in a stimulus detection task; this analysis provided clear evidence that the perceptual detection threshold fluctuated with the phase of the ongoing EEG signals (Busch et al. 2009). Another

study also suggested, by pooling of datasets from different designs, that pre-stimulus alpha phase has an impact on the behavioral outcome (VanRullen 2016).

In our study, we have employed an orthogonal design combining multi-modal attention and the prediction task. We tested their relation with an ANOVA in both induced power and coupling measures. In both tests, we did not find clusters reflecting significant interactions which suggests that the shifts of attention between two different modalities did not, or at least not significantly, interfere with the predicting process.

Prediction-related network dynamics

A role of neural oscillatory activity in predictive processing has been postulated early on and is supported by work on unimodal processing (for review, see e.g. Engel et al 2001; Engel and Fries 2010; Arnal and Giraud 2012). It has been suggested that predictions regarding stimulus identity or stimulus timing may involve oscillatory mechanisms in different frequency ranges. While the latter may be associated with delta and theta band oscillations, the former might be mediated by beta and gamma band interactions (Engel and Fries 2010; Arnal and Giraud 2012). So far, only relatively few studies have addressed the relation between oscillatory activity and predictions in the context of crossmodal processing (for review, see Arnal and Giraud 2012; Jessen and Kotz 2013). One of the earliest studies was carried out by Widmann and coworkers (2007) who used a matching paradigm in which visual score-like symbols predicted corresponding sound patterns. Compared to incongruent sounds, expected sounds evoked higher gamma-band activity. Van Ede et al (2010) presented subjects with a repeated series of concurrent tactile and auditory stimuli in which rare deviant stimuli occurred in one of the two modalities. Interestingly, prediction-related modulation of beta-band oscillations was observed even for unattended tactile stimuli. This modulation was enhanced during states of attentive expectation. Arnal et al (2011) investigated violation of multisensory predictions in speech perception and demonstrated that correct crossmodal predictions were associated with delta

oscillations, whereas prediction errors were predominantly accompanied by beta- and gamma-band activity. Another recent study on crossmodal effects in speech perception (Biau et al 2015) has observed a modulation of theta-oscillations during auditory speech processing by concomitant visual input from observation of gestures. Low-frequency oscillations have also been found to be associated with temporal prediction in a study where a visual color change predicted the occurrence of an auditory stimulus (van Wassenhove and Grzeczkowski 2015). We have shown that visually induced expectation of pain stimuli is associated with an anticipatory reduction of alpha-band activity (Höfle et al 2013). Altogether, only few studies are currently available that have addressed changes in oscillatory activity during crossmodal prediction.

To our knowledge, this study presents the first investigation of neural oscillations and their coupling in networks underlying crossmodal prediction of stimulus sequences. In the analysis of task-related power changes, we observed significant differences between the prediction task and the memory control condition. The cluster reflecting these differences was mainly found in the beta range, comprising mainly bilateral superior temporal gyri, sensorimotor cortices, insulae, cingulate and frontal cortices. In addition, there was an alpha component during sampling of the sequence but not in the predicting phase of the trials. This alpha component, which was mainly observed in visual areas, was more pronounced for the memory control task. This suggests that functional inhibition, as indicated by the alpha oscillations, was stronger during the sampling phase in the memory task. In this control condition, the information present in the sampling segments was task-irrelevant, whereas it was vital to construct the sequence in the prediction condition. Thus, it was not surprising that functional inhibition occurred for visual items in the memory task, while inhibition of visual processing was released during the sampling phase in the prediction task. Beta power was stronger during the memory task than during the prediction task, in line with the proposal that beta band oscillations reflect the maintenance of the task-relevant information in a cognitive state (Engel and Fries 2010). In the

prediction task, the participants needed to constantly update the sequence information after presentation of an item and predict the upcoming sensory inputs in the interval before the occurrence of the subsequent item; this cognitive updating might relate to the observed reduction of beta power.

In the same network, we also observed power difference in the responding segments, where subjects had to indicate whether the last item preserved or violated the acquired sequence. These data suggest a beta power decrease upon violation of the sequence, with a subsequent beta rebound. This dynamics of beta power shows similarities to changes associated with prediction errors, as proposed in recent predictive coding frameworks (Arnal and Giraud 2012; Bastos et al 2012).

In addition to the task-related power differences, we also observed functional connectivity clusters defined by coupling differences between the prediction and the memory task. All clusters were similar in their spatial distribution, and they also shared a large proportion of the involved brain regions with the cluster obtained by comparing power difference. This suggests that these clusters, although formally derived in separate analysis steps, reflect prediction-related dynamics in the same cortical network. This network comprised mainly superior temporal gyri, sensorimotor cortices, insulae, cingulate and frontal cortices. in both hemispheres. An interesting observation was that motor and pre-motor areas were quite prominent in the prediction network. This is in line with previous fMRI studies that have demonstrated a critical involvement of motor and premotor regions in prediction tasks (Bubic et al. 2010; Arnal and Giraud 2012). In this network, coupling in the theta-alpha range was stronger for the prediction task compared to the memory task during both the sampling and the prediction phase of the trials, suggesting that coupling at low frequencies was essential for the prediction process. Interestingly, in the same network, coupling also occurred in the gamma frequency range, which was stronger for the memory condition. A possibility is that, in the memory task, the comparison of low-level visual or auditory feature promoted gamma coupling,

which has been related to feedforward transfer of sensory information (Engel and Fries 2010; Bastos et al 2015). In contrast, in the prediction task the information on the 5-item sequence needed to be constantly retrieved and updated, which could require more feedback communication that may primarily be conveyed by coupling of low-frequency oscillations (Bastos et al. 2015; Fries 2015). Interestingly, changes in theta-band coupling were also related to the processing of sequence preservation vs. violation in the response segments of the prediction trials.

Oscillatory coupling has been proposed to mediate communication between remote areas by providing dynamic window of phase relations optimally geared to permit signal transmission (Engel et al. 2001; Fries 2005, 2015; Womelsdorf et al. 2007). This concept implies that behavioral performance should also be associated with changes in the phase relations in the network and more efficient at particular phases of coupling, but impaired at other phase relations between the recorded sites. We examined how the behavioral performance was tuned by the theta phase in the prediction network described above. In the prediction task, both accuracy and response times exhibited a sinusoidal tuning by the theta phase in the predicting phase of the trials. This effect was much weaker in the memory trials. In the prediction task, the participants needed to acquire the information on the sequence and, thus phase relations may have been more strongly employed to coordinate signals across the involved brain areas; in contrast, the memory task possibly did not demand the same amount of coordination across regions of the prediction network. Similar effects have been demonstrated in cellular recordings in animal studies (Womelsdorf et al 2007; Ni et al 2016). Our study is, to our best knowledge, the first demonstration of this sinusoidal tuning of behavior by the phase of oscillatory coupling at the level of whole-brain recordings.

Furthermore, we investigated whether power changes occurring during the prediction task also reflected the stimulus context, i.e., whether the preceding stimulus would have a contextual effect on the current one, and whether processing of the current item already predicted

processing of a subsequent item. Both was observed to be the case and to be reflected in changes of beta power, which was larger if the same item appeared before or after the current item, suggesting that the power difference related to differences in the storage and expectation of information in the network. The storage effect was more pronounced during the prediction than during the sampling phase, possibly reflecting that participants needed to actively access the remembered information of previous items to update prediction of the forthcoming events. The expectation effect could only be observed in the predicting phase – only when they knew the sequence. Interestingly, this effect mainly involved the motor, supplementary motor and premotor regions. This strongly support the idea that prediction is an active process, which relies on areas involved in action control even in the absence of an overt motor response (Engel et al. 2001; Bubic et al 2010). The results of our phase-amplitude coupling analysis for the theta and beta signals in the prediction network suggest that beta power, through coupling to the local theta phase, might dynamically modulate the information transfer through the long-range theta phase-phase coupling. Similar phase-amplitude coupling between theta and beta signals has been observed in previous studies, e.g., in visual working memory paradigms (Daume et al 2017).

Taken together, the power and coupling changes observed for the prediction network suggest the following possible scenario underlying processing of sequence information in our task setting: In the sampling phase, encoding of the presented items may be differentially modulated by changes in beta power already in sensory regions, depending on the task (prediction vs. memory) and on the contextual information. In the prediction phase of the trials, incoming new information is compared with the stored information on the sequence, which induces further modulation of local beta power. Local phase-amplitude coupling between beta and theta signals may then shape the transfer of the task-relevant information by long-range theta band coupling in the prediction network. After transfer of prediction-related information, though theta phase

coupling, to target motor-related areas, the information might be recoded again through theta phase to beta power coupling, which then might have an impact on response selection.

In conclusion, we have investigated the neural mechanism of crossmodal prediction, combined with intermodal shifts of attention. Our data provide direct evidence that oscillations and phase coupling across remote regions play an important role in prediction. In particular, our results support the notion that the phase of oscillatory signals can modulate behavioral performance, in both the attention network and the prediction network involved in our task setting. These findings substantially corroborate and extend previous studies on the functional role of oscillations in gating of attention and communication through coherence.

Acknowledgements

We thank Guido Nolte for methodological discussions and Till Schneider for valuable comments on the experimental design. We thank Roger Zimmermann, Christiane Reißmann and Karin Deazle for aids in MEG data collection and participant recruitment. This work was supported by grants from the Deutsche Forschungsgemeinschaft (grant number SFB TRR 169/B1 and SFB 936/A3 to A.K.E.).

References

Arnal LH, Doelling KB, Poeppel D. 2015. Delta-beta coupled oscillations underlie temporal prediction accuracy. Cereb Cortex. 25:3077-3085.

Arnal LH, Giraud AL. 2012. Cortical oscillations and sensory predictions. Trends Cogn Sci. 16:390-398.

Arnal LH, Wyart V, Giraud AL. 2011. Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. Nat Neurosci. 14:797-801.

Bastos A, Usrey W, Adams R, Mangun G, Fries P, Friston K. 2012. Canonical microcircuits for predictive coding. Neuron. 76:695-711.

Bastos AM, Vezoli J, Bosman CA, Schoffelen J-M, Oostenveld R, Dowdall JR, de Weerd P, Kennedy H, Fries P. 2015. Visual areas exert feedforward and feedback influences through distinct frequency channels. Neuron. 85:390-401.

Biau E, Torralba M, Fuentemilla L, de Diego Balaguer R, Soto-Faraco S. 2015. Speaker's hand gestures modulate speech perception through phase resetting of ongoing neural oscillations. Cortex. 68:76-85.

Brainard DH. 1997. The psychophysics toolbox. Spatial Vision. 10:433-436.

Bubic A, von Cramon DY, Schubotz RI. 2010. Prediction, cognition and the brain. Front Hum Neurosci. 4:25.

Busch NA, Dubois J, VanRullen R. 2009. The phase of ongoing EEG oscillations predicts Visual Perception. J Neurosci. 29:7869-7876.

Dale AM, Fischl B, Sereno MI. 1999. Cortical surface-based analysis. I. Segmentation and surface reconstruction. Neuroimage. 9:179-194.

Daume J, Gruber T, Engel AK, Friese U. 2017. Phase-amplitude coupling and long-range phase synchronization reveal frontotemporal interactions during visual working memory. J Neurosci. 37:313-322.

Doesburg SM, Emberson LL, Rahi A, Cameron D, Ward LM. 2008. Asynchrony from synchrony: long-range gamma-band neural synchrony accompanies perception of audiovisual speech asynchrony. Exp Brain Res. 185:11-20.

Engel AK, Fries P. 2010. Beta-band oscillations--signalling the status quo? Curr Opin Neurobiol. 20:156-165.

Engel AK, Fries P, Singer W. 2001. Dynamic predictions: oscillations and synchrony in top-down processing. Nat Rev Neurosci. 2:704-716.

Foxe JJ, Snyder AC. 2011. The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. Front Psychol. 2:154.

Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Segonne, F., Salat, D.H., Busa, E., Seidman, L.J., Goldstein, J., Kennedy, D., et al. (2004). Automatically parcellating the human cerebral cortex. Cereb Cortex 14, 11-22.

Fries P. 2005. A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. Trends Cogn Sci. 9:474-480.

Fries P. 2015. Rhythms for cognition: communication through coherence. Neuron. 88:220-235.

Friese U, Daume J, Göschl F, König P, Wang P, Engel AK. 2016. Oscillatory brain activity during multisensory attention reflects activation, disinhibition, and cognitive control. Sci Rep. 6:32775.

Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, Parkkonen L, Hamalainen MS. 2014. MNE software for processing MEG and EEG data. Neuroimage. 86:446-460.

Hipp JF, Engel AK, Siegel M. 2011. Oscillatory synchronization in large-scale cortical networks predicts perception. Neuron. 69:387-396.

Hipp JF, Siegel M. 2013. Dissociating neuronal gamma-band activity from cranial and ocular muscle activity in EEG. Front Hum Neurosci. 7:338.

Hipp JF, Siegel M. 2015. Accounting for linear transformations of EEG and MEG data in source

analysis. PloS One. 10:e0121048.

Höfle M, Pomper U, Hauck M, Engel AK, Senkowski D. 2013. Spectral signatures of viewing a needle approaching one' own body when anticipating pain. Eur J Neurosci. 7:3089-3098.

Hyvarinen A. 1999. Fast and robust fixed-point algorithms for independent component analysis. IEEE Trans Neural Netw. 10:626-634.

Jensen O, Mazaheri A. 2010. Shaping functional architecture by oscillatory alpha activity: gating by inhibition. Front Hum Neurosci. 4:186.

Jensen O, Gips B, Bergmann TO, Bonnefond M. 2014. Temporal coding organized by coupled alpha and gamma oscillations prioritize visual processing. Trends Neurosci. 37:357-369.

Jessen S, Kotz SA. 2013. On the role of crossmodal prediction in audiovisual emotion perception. Frontiers Hum Neurosci. 7:369.

Kopell N, Ermentrout GB, Whittington MA, Traub RD. 2000. Gamma rhythms and beta rhythms have different synchronization properties. Proc Natl Acad Sci USA. 97:1867-1872.

Kujala J, Gross J, Salmelin R. 2008. Localization of correlated network activity at the cortical level with MEG. Neuroimage. 39:1706-1720.

Lange J, Christian N, Schnitzler A. 2013. Audio-visual congruency alters power and coherence of oscillatory activity within and between cortical areas. Neuroimage. 79:111-120.

Leopold DA, Murayama Y, Logothetis NK. 2003. Very slow activity fluctuations in monkey visual cortex: implications for functional brain imaging. Cereb Cortex. 13:422-433.

Mazziotta JC, Toga AW, Evans A, Fox P, Lancaster J. 1995. A probabilistic atlas of the human brain: theory and rationale for its development. The International Consortium for Brain Mapping (ICBM). Neuroimage. 2:89-101.

Menon V, Uddin LQ. 2010. Saliency, switching, attention and control: a network model of insula function. Brain Struct Funct. 214:655-667.

Misselhorn J, Friese U, Engel AK. 2019. Frontal and parietal alpha oscillations reflect

attentional modulation of cross-modal matching. Sci Rep. 9: 5030.

Mitra PP, Pesaran B. 1999. Analysis of dynamic brain imaging data. Biophys J. 76:691-708.

Ni J, Wunderle T, Lewis CM, Desimone R, Diester I, Fries P. 2016. Gamma-rhythmic gain modulation. Neuron. 92:240-251.

Nichols TE, Holmes AP. 2002. Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum Brain Mapp. 15:1-25.

Nolte G. 2003. The magnetic lead field theorem in the quasi-static approximation and its use for magnetoencephalography forward calculation in realistic volume conductors. Phys Med Biol. 48:3637-3652.

Nolte G, Bai O, Wheaton L, Mari Z, Vorbach S, Hallett M. 2004. Identifying true brain interaction from EEG data using the imaginary part of coherency. Clin Neurophysiol. 115:2292-2307.

Oostenveld R, Fries P, Maris E, Schoffelen JM. 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intel Neurosc. 2011:156869.

Senkowski D, Schneider TR, Foxe JJ, Engel AK. 2008. Crossmodal binding through neural coherence: implications for multisensory processing. Trends Neurosci. 31:401-409.

Senkowski D, Hofle M, Engel AK. 2014. Crossmodal shaping of pain: a multisensory approach to nociception. Trends Cogn Sci. 18:319-327.

Stolk A, Todorovic A, Schoffelen JM, Oostenveld R. 2013. Online and offline tools for head movement compensation in MEG. Neuroimage. 68:39-48.

Talsma D, Senkowski D, Soto-Faraco S, Woldorff MG. 2010. The multifaceted interplay between attention and multisensory integration. Trends Cogn Sci. 14:400-410.

Tort AB, Komorowski R, Eichenbaum H, Kopell N. 2010. Measuring phase-amplitude coupling between neuronal oscillations of different frequencies. J Neurophysiol. 104:1195-1210.

Utevsky AV, Smith DV, Huettel SA. 2014. Precuneus is a functional core of the default-mode

network. J Neurosci. 34:932-940.

van Ede F, Jensen O, Maris E. 2010. Tactile expectation modulates pre-stimulus beta-band oscillations in human sensorimotor cortex. Neuroimage. 51:867-876.

van Driel J, Knapen T, van Es DM, Cohen MX. 2014. Interregional alpha-band synchrony supports temporal cross-modal integration. Neuroimage. 101:404-415.

Van Veen BD, van Drongelen W, Yuchtman M, Suzuki A. 1997. Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. IEEE Trans Biomed Eng. 44:867-880.

van Wassenhove V, Grzeczkowski L. 2015. Visual-induced expectations modulate auditory cortical responses. Frontiers Neurosci. 9:11.

VanRullen R. 2016. Perceptual cycles. Trends Cogn Sci. 20:723-735.

von Stein A, Sarnthein J. 2000. Different frequencies for different scales of cortical integration: from local gamma to long range alpha/theta synchronization. Intern J Psychophysiol. 38:301-313.

Wang P, Göschl F, Friese U, König P, Engel AK. 2019. Long-range functional coupling predicts performance: Oscillatory EEG networks in multisensory processing. Neuroimage. 196:114-125.

Widmann A, Gruber T, Kujala T, Tervaniemi M, Schröger E. 2007. Binding symbols and sounds: evidence from event-related oscillatory gamma-band activity. Cereb Cortex. 17: 2696-2702.

Womelsdorf T, Schoffelen JM, Oostenveld R, Singer W, Desimone R, Engel AK, Fries P. 2007. Modulation of neuronal interactions through neuronal synchronization. Science. 316:1609-1612.

Womelsdorf T, Valiante TA, Sahin NT, Miller KJ, Tiesinga P. 2014. Dynamic circuit motifs underlying rhythmic gain control, gating and integration. Nat Neurosci. 17:1031-1039.

Supplementary Figures

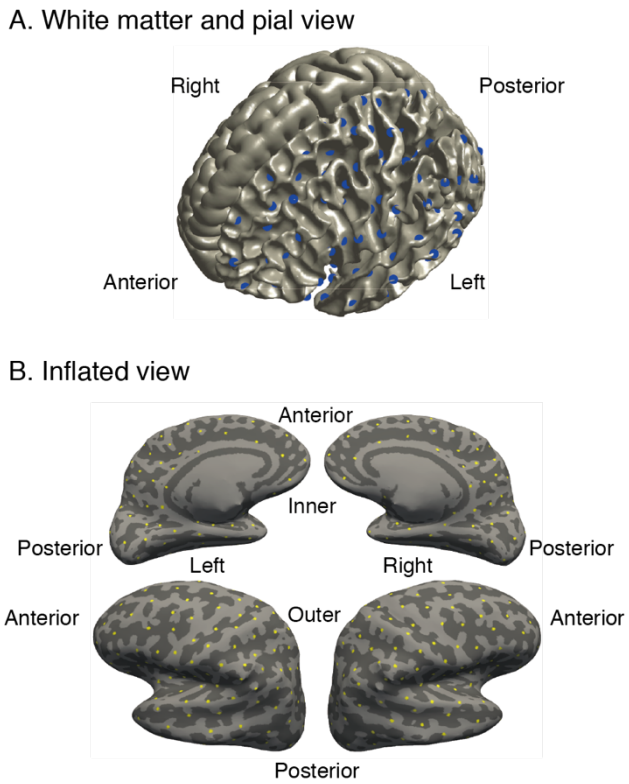

Figure S1: Definition of the source space. (A) White matter (left hemisphere) and pial (right hemisphere) surfaces of the template brain. The source locations were evenly distributed on the surface of the white matter, as indicated by the blue dots in the left hemisphere. (B) The inflated view of the template brain. The source locations are indicated by the yellow dots on the inflated surface.

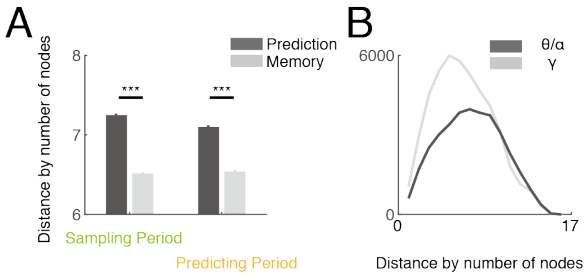

Figure S2: Path length dependency on oscillation frequencies. (A) Comparison of average path length as indicated by number of nodes, in the topological anatomy structure, for clusters that favorites prediction or memory tasks. (B) Distribution of path lengths for all connections in γ or θ/α frequency range.

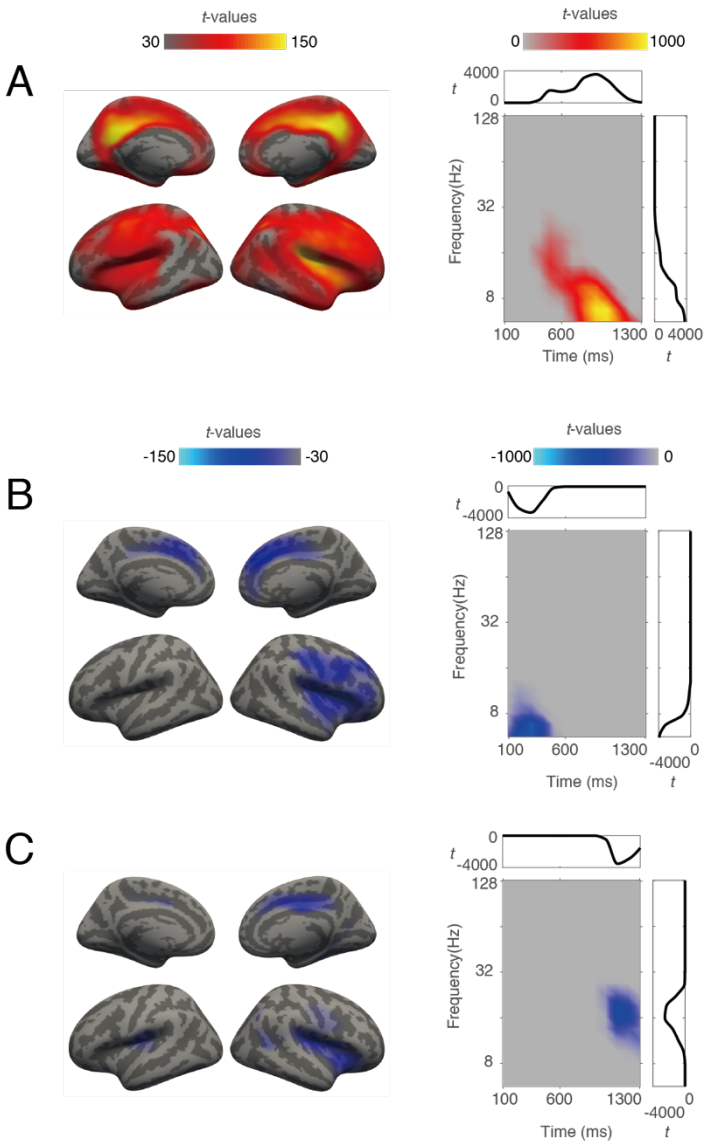

Figure S3: Cluster analysis of power during responding segments. (A - C). In each panel, both spatial (left) and time-frequency (right) distribution are illustrated. Color indicates accumulated *t* values. Side plots in box are integrated *t* values against time (top, x-axis: time in ms, y-axis: *t* value) or frequency (right, x-axis: *t* value, y-axis: frequency in Hz).

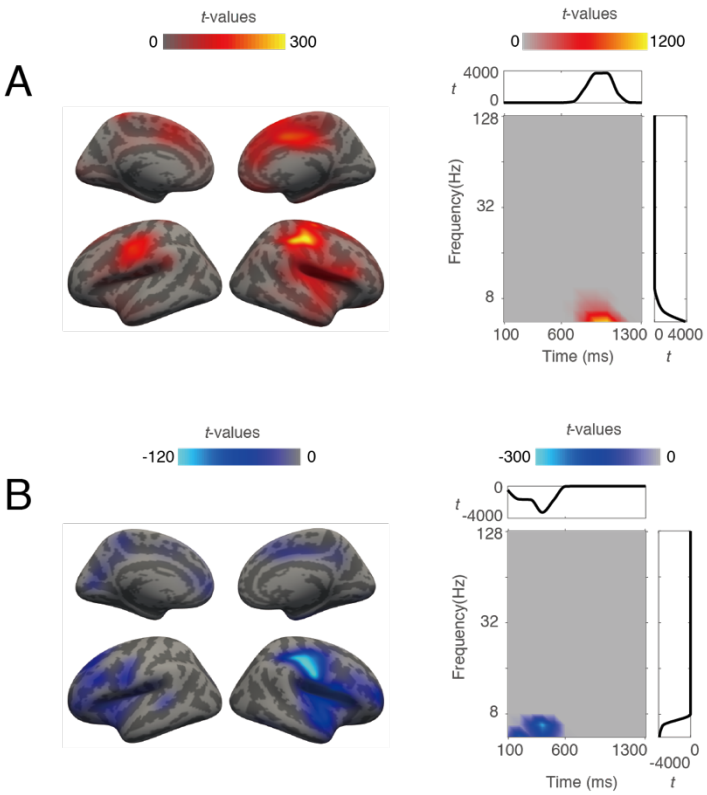

Figure S4: Cluster analysis of coupling during responding segments. In each panel, both spatial (left) and time-frequency (right) distribution are illustrated. Color indicates accumulated *t* values. Side plots in box are integrated *t* values against time (top, x-axis: time in ms, y-axis: *t* value) or frequency (right, x-axis: *t* value, y-axis: frequency in Hz).