Cross-modal interaction of Alpha Activity does not reflect inhibition of early sensory processing: A frequency tagging study using EEG and MEG

Brickwedde, M.[1,2,3], Limachya, R.[3], Markiewicz, R. [4], Sutton, E. [3], Postzich, C.[5], Shapiro, K. [3], Jensen, O[3]., and Mazaheri, A[3].

[1]Department of Child and Adolescent Psychiatry, Charité - Universitätsmedizin Berlin, Germany

[2]Physikalisch Technische Bundesanstalt, Berlin, Germany

[3]Centre for Human Brain Health, School of Psychology, University of Birmingham, Uk

[4]Istituto Italiano di Tecnologia, Genova, Italy

[5]Max-Planck Research Institute, Leipzig, Germany

Abstract

Selective attention involves prioritizing relevant sensory input while suppressing irrelevant stimuli. It has been proposed that oscillatory alpha-band activity (~10 Hz) aids this process by functionally inhibiting early sensory regions. However, recent studies have challenged this notion. Our EEG and MEG studies aimed to investigate whether alpha oscillations serve as a 'gatekeeper' for downstream signal transmission. We first observed these effects in an EEG study and then replicated them using MEG, which allowed us to localize the sources.

We employed a cross-modal paradigm where visual cues indicated whether upcoming targets required visual or auditory discrimination. To assess inhibition, we utilized frequency-tagging, simultaneously flickering the fixation cross at 36 Hz and playing amplitude-modulated white noise at 40 Hz during the cue-to-target interval.

Consistent with prior research, we observed an increase in posterior alpha activity following cues signalling auditory targets. However, remarkably, both visual and auditory frequency tagged responses amplified in anticipation of auditory targets, correlating with alpha activity amplitude. Our findings suggest that when attention shifts to auditory processing, the visual stream remains responsive and is not hindered by occipital alpha activity. This implies that alpha modulation does not solely regulate 'gain control' in early sensory areas but rather orchestrates signal transmission to later stages of the processing stream.

Introduction

In our daily life, we are often confronted with sensory information from many different sources, all at once. To operate effectively, we require selective attention, reconciling the tension between environmental inputs relevant for top-down goals and sensory information that may be perceptually salient but task-irrelevant. Previous research has suggested that oscillatory activity in the alpha range (~10 Hz) plays a mechanistic role in selective attention through functional inhibition of irrelevant cortices (see Fig. 1A; Foxe et al., 1998; Jensen & Mazaheri, 2010; Klimesch et al., 2007). The concept of functional inhibition refers to an area of the cortex being actively hindered to process input, which is distinctly different from idling, where a part of the cortex is not actively involved. Evidence supporting the functional inhibition framework of alpha modulation revealed an increase in alpha-power over task-irrelevant sensory cortices after the onset of cues indicating the spatial location (Kelly et al., 2006; Okazaki et al., 2014; Thut et al., 2006; Worden et al., 2000; Zumer et al., 2014) or specific modality of an upcoming target (Foxe et al., 1998; Fu et al., 2001; Mazaheri et al., 2014). Moreover, previous investigations have observed 'spontaneous fluctuations' of alpha power in sensory regions, particularly in the visual cortex, to be inversely related to discrimination ability (Ergenoglu et al., 2004; Van Dijk et al., 2008). Alpha inhibition is believed to be transmitted in a phasic manner, as phosphene perception as well as high-frequency and spiking activity vary in line with the alpha cycle (Dugué et al., 2011; Haegens et al., 2011; Spaak et al., 2012).

As recent evidence contradicted direct connection between alpha activity and sensory information processing in early visual cortex (Zhigalov & Jensen, 2020), the objective of the current study is to investigate whether alpha modulation is connected to 'gain control' in early sensory areas through modulation of excitability (Foxe & Snyder, 2011; Jensen & Mazaheri, 2010; Van Diepen et al., 2019) or whether inhibitory effects are exhibited at later stages of the processing stream (Yang et al., 2023; Zhigalov & Jensen, 2020; Zumer et al., 2014), gating

feedforward or feedback communication between sensory areas (Bauer et al., 2020; Haegens et al., 2015; Uemura et al., 2021).

To this end, we applied frequency-tagging, the rhythmic presentation of sensory stimuli, which elicits steady-state sensory evoked potentials or fields (SSEP/SSEF), consisting of rhythmic neuronal activity in the frequency of stimulation (Brickwedde et al., 2020; Colon et al., 2012; Dinse et al., 1998; Marzoll et al., 2018; Regan, 1982; Snyder, 1992; Stapells et al., 1984; Tobimatsu et al., 1999). The magnitude of SSEPs is attention-dependent (de Jong et al., 2010; Müller et al., 1998; Müller & Hillyard, 2000; Porcu et al., 2013; Saupe et al., 2009; Toffanin et al., 2009) even for frequencies too fast to perceive consciously (Brickwedde et al., 2022; Zhigalov et al., 2019). Their scalp topography reveals that SSEPs are most strongly observable over occipital areas for visual stimuli generated in the visual cortex, and over temporal (MEG) or fronto-to-central (EEG) areas for auditory stimuli generated in the auditory cortex (de Jong et al., 2010; Hari et al., 1989; Pantev et al., 1996; Regan, 1982). Their distinct response, localisation and attention-dependence provide an optimal tool to study sensory signal processing over time.

The aim of our initial EEG study was to directly investigate if the cue-induced modulation of alpha activity coincides with the suppression of frequency-tagging responses in task-irrelevant modalities. Based on previous studies, we utilized a cross-modal attention paradigm, in which symbolic visual cues signalled the target modality (visual or auditory) of an upcoming discrimination task (e.g., van Diepen et al., 2015). Here, we included an additional experimental manipulation in the form of frequency-tagging to assess the involvement of the auditory and visual systems in the *cue-to-target interval*. We were also interested in the relationship between alpha modulation, SSEPs, and attentional performance on a trial-by-trial basis.

In line with previous results, we hypothesized that signalling an upcoming auditory target would lead to increased alpha activity over visual occipital regions as well as increased SSEP

responses to auditory and decreased SSEP to visual stimuli as indexed by frequency-tagging. Furthermore, we aimed to explore whether there is a direct connection between alpha activity and frequency-tagging responses in other areas than primary visual cortex. In brief, while we observed the expected cue-induced alpha modulation in our initial EEG-study, the amplitude of auditory as well as *visual* SSEPs increased just prior to the onset of the auditory target.

To confirm our novel findings as well as analyse the cortical origin of the observed SSEP-changes, we replicated our study using MEG. Here we made slight adjustments to the experimental design by implementing a blocked design. The initial block involved no distractors, while the subsequent block consistently included a distractor. Furthermore, we modified the visual orientation task to align more closely with the auditory task.

The transition to a blocked design allows for a more controlled comparison between conditions with and without distractors, providing insights into how the presence of distractors influences the observed effects. Additionally, modifying the visual orientation task to align with the auditory task may enhance the comparability of the two modalities and potentially reveal more nuanced interactions between attentional processes and sensory processing across modalities. Here we not only replicated the observation that the amplitude of both *visual and auditory* SSEPs increased just prior to the onset of the auditory target, but we are also able to localise the sources of these activities.

Results

To assess audio-visual excitability in anticipation of either visual or auditory targets, we applied visual cues to signal the modality of the upcoming target (auditory, visual, or unspecified). In a three-second cue-to-target interval, we frequency-tagged the fixation cross at 36 Hz and played 40 Hz amplitude modulating white noise. Participants had to either discriminate between three different pitch sounds (auditory target) or three different Gabor patch orientations (visual

target). If the target modality was cued, 50% of the trials were accompanied by a random distractor from the target pool of the opposing sensory modality (see Fig. 1B). In our subsequent MEG-study, we slightly adjusted the experiment into a blocked design. Specifically, the first block contained no distractors while in the second block, a distractor was always present. Additionally, we exchanged the visual orientation task to be more in line with the auditory task. Participants now had to distinguish between three different frequencies of the Gabor patch stripes. Prior to the experiment, difficulty of pitch sounds, and Gabor patch frequency were calibrated for each individual, ascertaining a success rate between 55% to 75%.

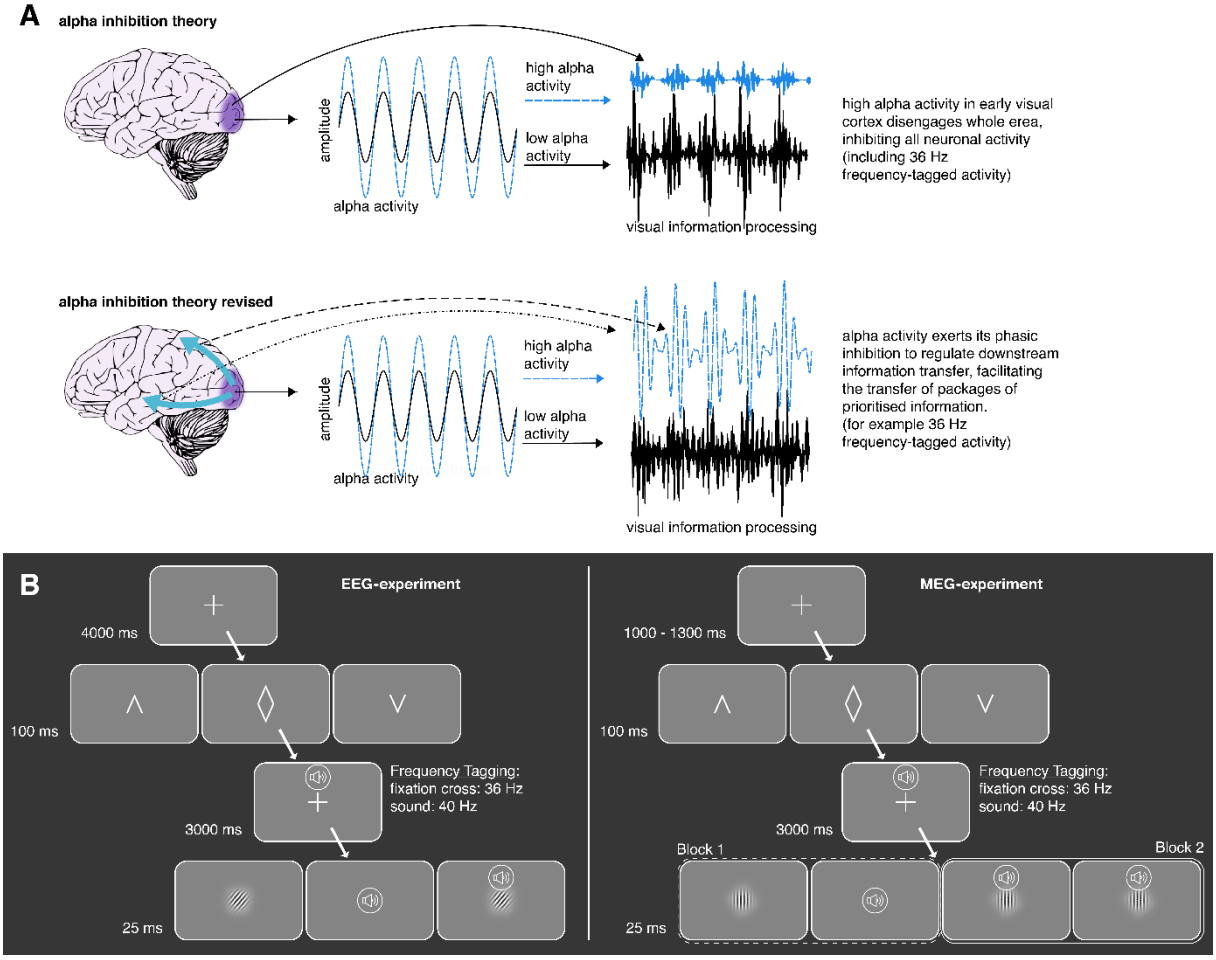


Figure 1. Illustration the alpha inhibition theory. A, the alpha inhibition theory suggests that alpha inhibits sensory information processing in a phasic manner. If alpha activity is high, it suggests that the whole area is inhibited and thereby disengaged (Foxe et al., 1998; Jensen & Mazaheri, 2010; Klimesch et al., 2007). We propose a revision of this theory, whereby alpha activity exerts its phasic inhibition to regulate downstream information transfer, creating enhanced signal packages of prioritised information (see also Yang et al., 2023; Zhigalov & Jensen, 2020; Zumer et al., 2014). **B, Illustration of the cross-model discrimination task.** LEFT: in the EEG-experiment, trials were separated by a 4 s

interval, in which a fixation cross was displayed. A brief central presentation of the cue (100 ms) initiated the trial, signalizing the target modality (see figure above from left to right: auditory, unspecified, visual). In the cue-to-target interval, the fixation cross was frequency-tagged at 36 Hz. At the same time, a sound was displayed over headphones, which was frequency-tagged at 40 Hz. Both tones and fixation cross contained no task-relevant information. The target consisting either of a Gabor patch or a tone was presented for 25 ms, Participants had to differentiate between 3 different Gabor rotations or tone pitches, respectively. In 50% of auditory and visually cued trials, a distractor in form of a random pitch or rotation of the un-cued modality was presented alongside the target. RIGHT: The MEG-experiment followed an almost identical setup. This time, trials were separated by a 1 s interval followed by a random jitter of 0 to 300 ms. The visual task was adjusted to now require discrimination between 3 different Gabor patch frequencies for the visual task. Lastly, a blocked design was incorporated where in block 1, no distractors were presented while in block 2, a random distractor from the stimulus pool of the non-target modality was always presented.

Behavioural performance

In the EEG-study, we found that accuracy differed significantly between conditions ($F_{(5,105)} = 44.16$; $p < .001$). Overall, participants were significantly less accurate in the auditory discrimination task ('overall auditory', *M* = 79% correct, *SD* = 11.7) than in the visual discrimination task ('overall visual', *M* = 97%, *SD* = 2.3; see Fig. 2A), with the worst performance occurring when auditory targets were paired with visual distractors (auditory +: *M* = 74% correct, *SD* = 13). Our adjustments in the MEG-study streamlined performances to be more in line between auditory and visual conditions, especially in the second block ($F_{(3,75)} = 10.26$; $p < .01$). In block 1, participants were significantly more accurate in the auditory ('block 1 auditory', *M* = 84% correct, *SD* = 10.03) compared to the visual task ('block 1 visual', *M* = 70%, *SD* = 10.61; see Fig. 2C). However, in block 2, there was no observable difference between visual and auditory task accuracy ('block 2 auditory', *M* = 77% correct, *SD* = 10.03; ('block 2 visual', *M* = 78%, *SD* = 13.15).

Reaction times yielded a similar pattern, with auditory reactions (overall auditory: *M* = 662 ms, *SD* = 136) being slower than for the visual task (overall visual: *M* = 597 ms, *SD* = 130; main effect over all conditions: $F_{(5,105)} = 27.47$; $p < .001$; see Fig. 2B). This was mostly driven by slow responses to auditory targets paired with distractors (auditory +: *M* = 723 ms). In the

MEG-study, reaction times were comparable between conditions (overall: $M = 661$ ms, $SD = 76$), with the exception of responses to visual targets in block 1 (block 1 visual: $M = 726$ ms, $SD = 85$), which were significantly slower than all other reaction times (main effect condition: $F_{(3,75)} = 18.13$; $p < .001$; see Fig. 2D).

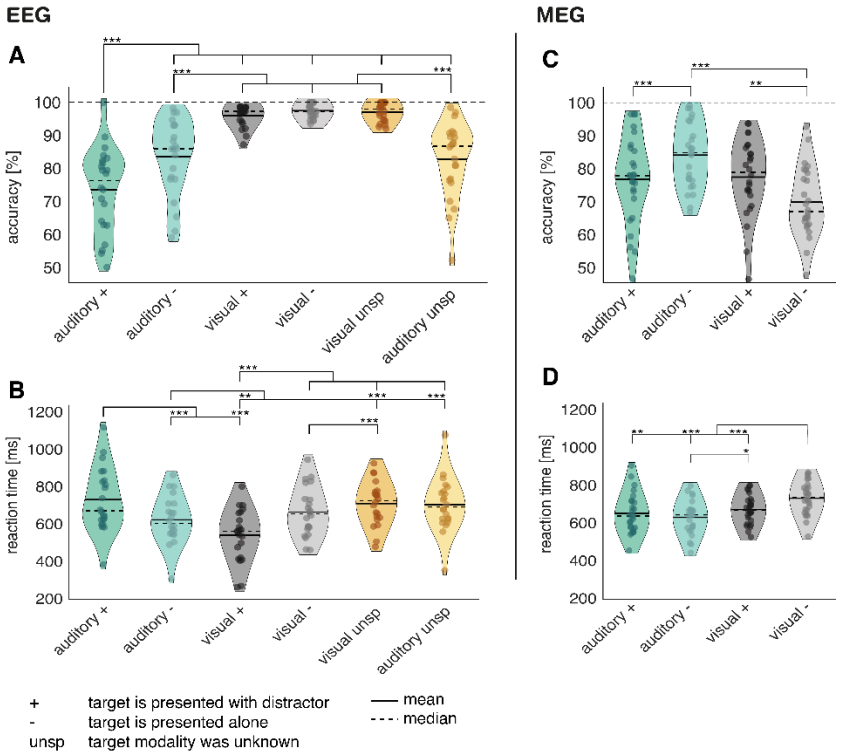


Figure 2. Analysis of task accuracy and reaction time indicates increased difficulty of auditory targets in the EEG study, and comparable difficulties in the MEG study. **A**, Task accuracy compared between all 6 experimental conditions reveals a drop in accuracy for responses to auditory targets. **B**, reaction times of correct trials compared between all 6 experimental conditions. The slowest reaction times are observable following auditory targets alongside visual distractors. **C**, Task accuracy differences were only observable in the first block without distractors. **D**, reaction times to visual targets in the first block were strongly decreased compared to all other conditions. In the second block, no significant difference in reaction times was observable. EEG Study: N = 22; MEG-Study: N = 27; *** sig < .001; ** sig. < .01; * sig. < .05;

Cues and distractors were behaviourally relevant, as both attentional benefit and distractor cost were observable in our data (see suppl Fig. 1). Interestingly, auditory distractors reduced reaction times to visual targets, which could be explained by a generally faster processing of auditory targets (Jain et al., 2015), possibly probing faster responses in visual tasks (Naue et

al., 2011). In the MEG-study, the distractor cost could not be differentiated from learning effects, due to the blocked design. Nonetheless, the previously observed pattern of the EEG study was replicated.

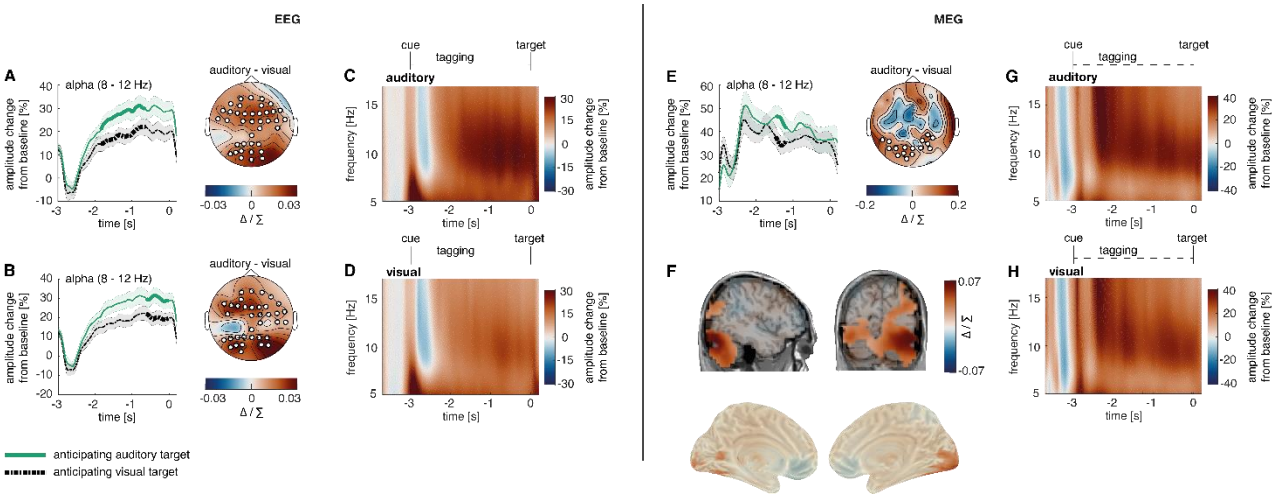


Figure 3. Post-cue modality specific early visual modulation of alpha power in anticipation of an auditory versus a visual target A-B, The time course of post-cue alpha power. Cluster permutation analysis resulted in two condition effects, both indicating heightened alpha activity when expecting an auditory compared to a visual target (C: $p < .01$; D: $p < .01$). **C-D**, Time-frequency representation of power in the cue-to-target interval. A greater increase in alpha power was observed when expecting an auditory target (average over significant electrodes for the condition difference in A). **E**, The time course of post-cue alpha power. Cluster permutation analysis resulted in a condition effects, indicating heightened alpha activity when expecting an auditory compared to a visual target ($p = .034$). **F**, source localization of the condition difference between expecting an auditory versus a visual target, revealing a significant cluster in early visual areas with stronger effects on the right hemisphere ($p < .01$). **G-H**, Time-frequency representation of power in the cue-to-target interval. A greater increase in alpha power was observed when expecting an auditory target (average over electrodes that showed maximal condition difference in E). **A, B, E,** Cluster electrodes are marked in white. Shading represents standard error from the mean; Δ / ∑ represents (a-b)/(a+b) normalization.

Cross-modal cues differentially modulated pre-target alpha activity

We conducted a time-frequency analysis of power in the cue-to-target interval and found a stronger amplitude increase from baseline for auditory compared to visual target conditions starting around 2 s before target onset (Fig. 3A-B). We also calculated the time course of alpha power changes using the Hilbert-transformation (8 – 12 Hz, Fig. 3C-D). Consistent with previous work (Mazaheri et al., 2014; van Diepen & Mazaheri, 2017), cluster permutation analysis conducted using the last two seconds before target onset revealed two clusters of difference in

alpha power when expecting an auditory compared to a visual target. These effects corresponded to clusters extending from -1.84 to – 0.64 s (p = .004) and -0.62 to 0s (p = .005). We did not find significant effects for the ambivalent condition, which we then excluded from further analyses (see suppl Fig. 2 for the data in this condition). In the MEG-experiment, in line with previous studies (van Diepen & Mazaheri, 2017), condition differences in alpha activity were only significant in block 2, where distractors in a different modality were presented. Therefore, we performed the same analyses as previously described for the EEG-study only for the second half of the experiment (see Fig. 3E-G), revealing a significant cluster from -1.47 to -1.18 s (p = .034). Applying roughly the same time-window (-1.5 and-2 s), we conducted a source localization, contrasting the two conditions with cluster permutation analysis. We found that condition differences were located in early visual areas with a stronger effect on the right compared to the left hemisphere (p < .01; peak spm coordinates: 41 -82 -19 mm, in the right lingual gyrus see Fig. 3H).

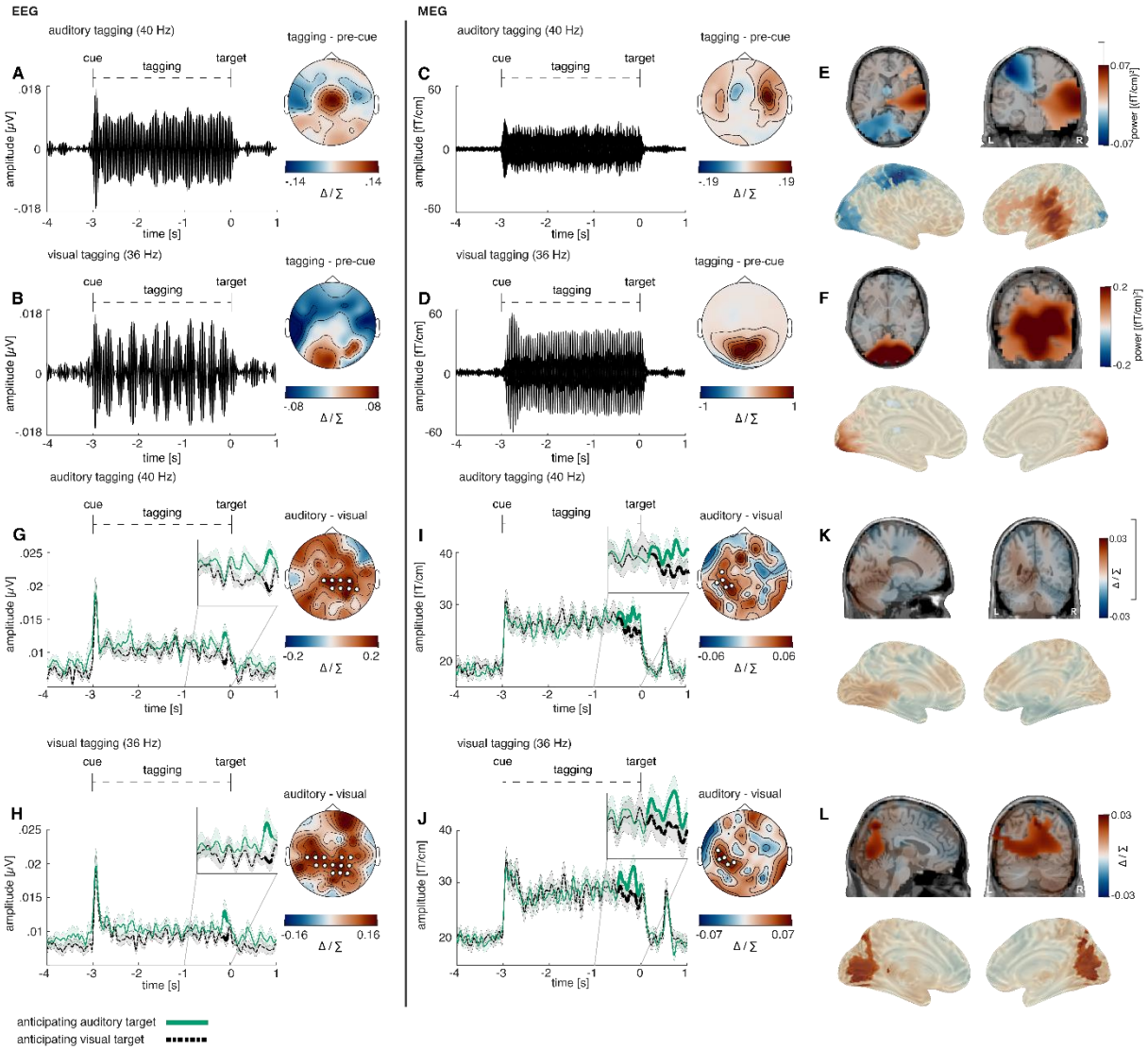


Figure 4. Increase in amplitude of both visual and auditory frequency tagged responses when anticipating visual or auditory targets Event-related potentials and scalp topographies reveal distinct modality specific responses at the tagged frequencies. **A)** auditory steady-state evoked potential (ASSEP) averaged over 6 central electrodes displaying the highest 40 Hz power (Fz, FC1, FC2, F11, F2, FCz). **B**, visual steady-state evoked potential (VSSEP) averaged over 4 occipital electrodes displaying the highest 36 Hz power (POz, O1, O2, Oz **C**, auditory steady-state evoked fields (ASSEF) averaged over 20 temporal sensors displaying the highest 40 Hz power (10 right, 10 left). **D**, visual steady-state evoked fields (VSSEF) averaged over 10 occipital sensors, displaying the highest 36 Hz power. **E**, ASSEF source localization revealed a significant positive cluster in the right-hemispheric early auditory cortex ($p < .001$). **F**, VSSEF source localization revealed a significant positive cluster in the early visual cortex ($p < .001$). **G,I**, In both the EEG and the MEG study, the Hilbert-envelope of the 40 Hz ASSEP/ASSEF reveals an increase shortly before target onset when anticipating an auditory compared to a visual target (EEG: $p = .041$; MEG: $p = .043$); **H,J**), The Hilbert-envelope of the 36 Hz VSSEP/VSSEF likewise reveals an increase shortly before target onset when anticipating an auditory compared to a visual target, both in the EEG as well as the MEG study (EEG: $p = .014$; MEG: $p = .019$).**K**, Condition differences in the 40 Hz ASSEF response did not reach significance in sensor space. **L**, Condition differences in the 36 Hz VSSEF response were significant over several areas of the visual stream, including most strongly the medial occipital cortex, the calcarine fissure, and the precuneus ($p = .047$); *note*: Cluster electrodes are marked in white. Shading represents standard error from the mean. Δ / ∑ represents (a-b)/(a+b) normalization.

Cross-modal cues increased the amplitude of the frequency-tagged responses across both modalities

To assess the temporal development of frequency-tagging responses, steady-state potentials were calculated using data band-pass filtered around the tagging frequency. Neuronal responses to the 40 Hz auditory tagging were strongest over central areas (Fig. 4A) and 36 Hz responses were strongest over occipital areas (Fig. 4B). In accordance with the EEG-data, neuronal responses to 40 Hz auditory tagging measured with MEG were strongest over temporal areas (Fig. 4C) and 36 Hz responses were strongest over occipital areas (Fig. 4D). As expected, the auditory tagging response originated from the right-hemispheric early auditory cortex (cluster significance: $p < .01$; peak spm coordinates: 69 -22 5 mm, in the right superior temporal gyrus; see Fig. 4E) and the visual tagging response originated from the early visual cortex (cluster significance: $p < .001$; peak spm coordinates: 19 -105- 11 mm, in the right lingual gyrus; see Fig. 4F). Additionally, there was a significant reduced 40 Hz activity in the left-hemispheric visual-to-central cortex (cluster significance: $p < .01$).

To assess the differences between conditions, the Hilbert envelope of the steady-state potentials was analysed using cluster permutation analyses. When expecting an auditory target, the auditory 40 Hz frequency-tagging response was larger shortly before target onset (see Fig. 4G ; -0.15 to -0.08 s; $p = .041$).

Surprisingly, the visual 36 Hz frequency-tagging response was likewise increased shortly before expecting an auditory compared to a visual target (see Fig. 4H; -0.16 to -0.06 s; $p = .014$). For both visual 36 Hz and auditory 40 Hz frequency-tagging responses, condition differences appeared strongest over mid-parietal regions. Applying the same analysis to MEG data over the last 500 ms before target onset replicated the EEG-results (see Fig. 4I-J; auditory target: $p = .043$; visual target: $p = .019$). As such, frequency tagging responses might reflect effort affecting the vigilance of the sensory system rather than the sensory-specific allocation of attention.

Source localization confirmed the condition difference in 36 Hz activity to originate from later stages of the processing stream, encompassing a wide range of areas, most strongly the medial occipital cortex, the calcarine fissure and the precuneus (cluster significance: $p = .047$; peak spm coordinates: 3 -66 44 mm, in the left and right precuneus; see Fig. 4L). In source space, the effect was not significant for 40 Hz activity (p = .11; peak spm coordinates: -9 -39 0 mm, in the left precuneus; see Fig. 4K).

Alpha power was positively correlated with amplitude of frequency tagged responses

Following the observation of condition differences in alpha activity and frequency-tagging responses, we were further interested in exploring whether these responses were connected. Accordingly, we conducted trial-by-trial correlations using alpha condition differences and their electrode positions as seed, which was correlated with frequency-tagging signals over all electrodes. Multiple comparison correction was applied by testing the correlation matrix against a zero-correlation matrix with a cluster permutation approach. A positive correlation was observed over right parietal-to-occipital areas between the late alpha cluster activity and both 40 Hz ($p = .009$) and 36 Hz ($p = .004$) frequency-tagging responses when expecting a visual target (see Fig. 5.A-B). This result is further illustrated by a median split analysis between trials with high and low alpha power for each participant. It was highly significant for the visual 36 Hz response (Fig. 5A, middle columns, $p = .033$; $t_{(19)} = 2.29$; $BF_{(10)} = 1.91$) but did not reach significance for the visual 40 Hz response (Fig. 5B, middle column; $p = 0.20$; $t_{(19)} = 1.32$; $BF_{(10)} = 0.49$). Additionally, we averaged the correlation coefficient of each participant and calculated a one-sample t-test against 0. Both tests indicate strong correlations with alpha activity for both 40 Hz (Fig. 5B, right column; $p < .001$; $t_{(19)} = 4.95$; $BF_{(10)} = 306.93$) and 36 Hz activity (Figure 6A, right column; $p < .01$; $t_{(19)} = 3.66$; $BF_{(10)} = 23.57$). Applying the same analysis to the last 500 ms before target onset, we could replicate these results in our MEG data (see Fig. 6F-G; expecting a visual target, 36 Hz response: cluster significance: $p < .01$; median split: $p < .001$; $t_{(24)} = 4.33$; $BF_{(10)} = 127$; t-test: $p < .001$; $t_{(24)} = 5.33$; $BF_{(10)} = 1272$; expecting a visual target,

40 Hz response: cluster significance: $p < .001$; median split: $p < .001$; $t_{(24)} = 7.05$; $BF_{(10)} = 59443$; t-test: $p < .001$; $t_{(24)} = 6.75$; $BF_{(10)} = 30515$).

The same positive correlation with alpha activity was found when expecting an auditory target only for 36 Hz activity ($p = .031$), but not for 40 Hz activity (see Fig. 6D). A median split between high and low alpha activity ($p = .005$; $t_{(19)} = 3.14$; $BF_{(10)} = 8.53$) and correlation coefficients ($p = .002$; $t_{(19)} = 3.52$; $BF_{(10)} = 17.76$) provided moderate to strong evidence for this effect. In our MEG dataset, alpha activity 500 ms before auditory target onset correlated with both 36 Hz activity (see Fig6. I-J; cluster significance: $p < .001$; median split: $p = .001$; $t_{(23)} = 3.62$; $BF_{(10)} = 25.51$; t-test: $p < .001$; $t_{(23)} = 4.60$; $BF_{(10)} = 216$) and 40 Hz activity (cluster significance: $p = .005$; median split: $p < .001$; $t_{(25)} = 3.75$; $BF_{(10)} = 36.40$; t-test: $p < .001$; $t_{(25)} = 4.06$; $BF_{(10)} = 73.61$).

Additionally, we compared how correlation coefficients between alpha activity and frequency-tagging differed when anticipating an auditory versus a visual target. Multiple comparison correction was applied with cluster permutation analysis. Interestingly, an interaction between the strength of the correlation associating alpha and 36 Hz activity and condition became apparent ($p = .044$; *see* Fig. 6E) and was observed most strongly over right-central electrodes. Comparing the correlation coefficients of participants over this cluster revealed a strong effect and even a change of direction in the correlation ($p < .001$; $t_{(21)} = -4.76$; $BF_{(10)} = 259.97$). Particularly, when expecting a visual target, there was a negative correlation between 36 Hz and alpha activity, which turned positive when expecting an auditory target. Both correlations also differed significantly from 0 (expecting a visual target: $p < .001$; $t_{(21)} = -3.87$; $BF_{(10)} = 39.62$; expecting an auditory target: $p = .020$; $t_{(21)} = 2.53$; $BF_{(10)} = 2.83$). In contrast to the previously observed positive correlation between alpha activity and 36 Hz activity, the significant electrode cluster was located more ventrally. This effect could possibly hint at

dynamic adaptability of oscillatory alpha effects on later processing stages. However, this could not be replicated in our MEG dataset.

It is further noteworthy that the correlation between alpha activity and 36 Hz frequency-tagging response when expecting a visual target, was also present when using the early alpha cluster as seed (~ 1200 to 400 ms before target onset, see Fig. 6C), in which case alpha activity preceded the 36 Hz activity (p = .016*)*. For this correlation, the median split between high and low alpha trials did not reach significance (p = .11; $t_{(20)}$ = 1.69; $BF_{(10)}$ =0.76). Testing correlation coefficients against 0 again revealed a significant effect (p = .003; $t_{(20)}$ = 3.39; $BF_{(10)}$ =14.22). We could confirm these findings in our MEG-dataset, revealing a significant correlation between alpha activity during last 1 to 1.5 s before target onset and 36 Hz frequency-tagging response during the last 500 ms prior to an auditory target (see Fig. 6K; cluster significance: p = .01; median split: p < .001; $t_{(24)}$ = 4.66; $BF_{(10)}$ = 271; t-test: p < .001; $t_{(24)}$ = 5.66; $BF_{(10)}$ = 2688). The same alpha activity correlated with 40 Hz activity during the last 500 ms prior to a visual target (see Fig. 6H; cluster significance: p = .002; median split: p = .002; $t_{(23)}$ = 3.41; $BF_{(10)}$ = 16.38; t-test: p < .001; $t_{(23)}$ = 5.57; $BF_{(10)}$ = 1901).

Lastly, both alpha activity as well as 36 Hz frequency-tagging 500 ms before target onset activity correlated negatively with reaction time on a trial-by-trial basis, indicating faster reaction times in trials with higher pre-stimulus activity (alpha: p = .037; median split: p = .013; $t_{(25)}$ = -2.67; $BF_{(10)}$ = 3.78; t-test: p < .01; $t_{(25)}$ = -3.34; $BF_{(10)}$ = 14.84; 36 Hz: p = .002; median split: p = .004; $t_{(25)}$ = -3.20; $BF_{(10)}$ = 8.98; t-test: p < .01; $t_{(25)}$ = -3.46; $BF_{(10)}$ = 19.12. See suppl Fig. 3-4).

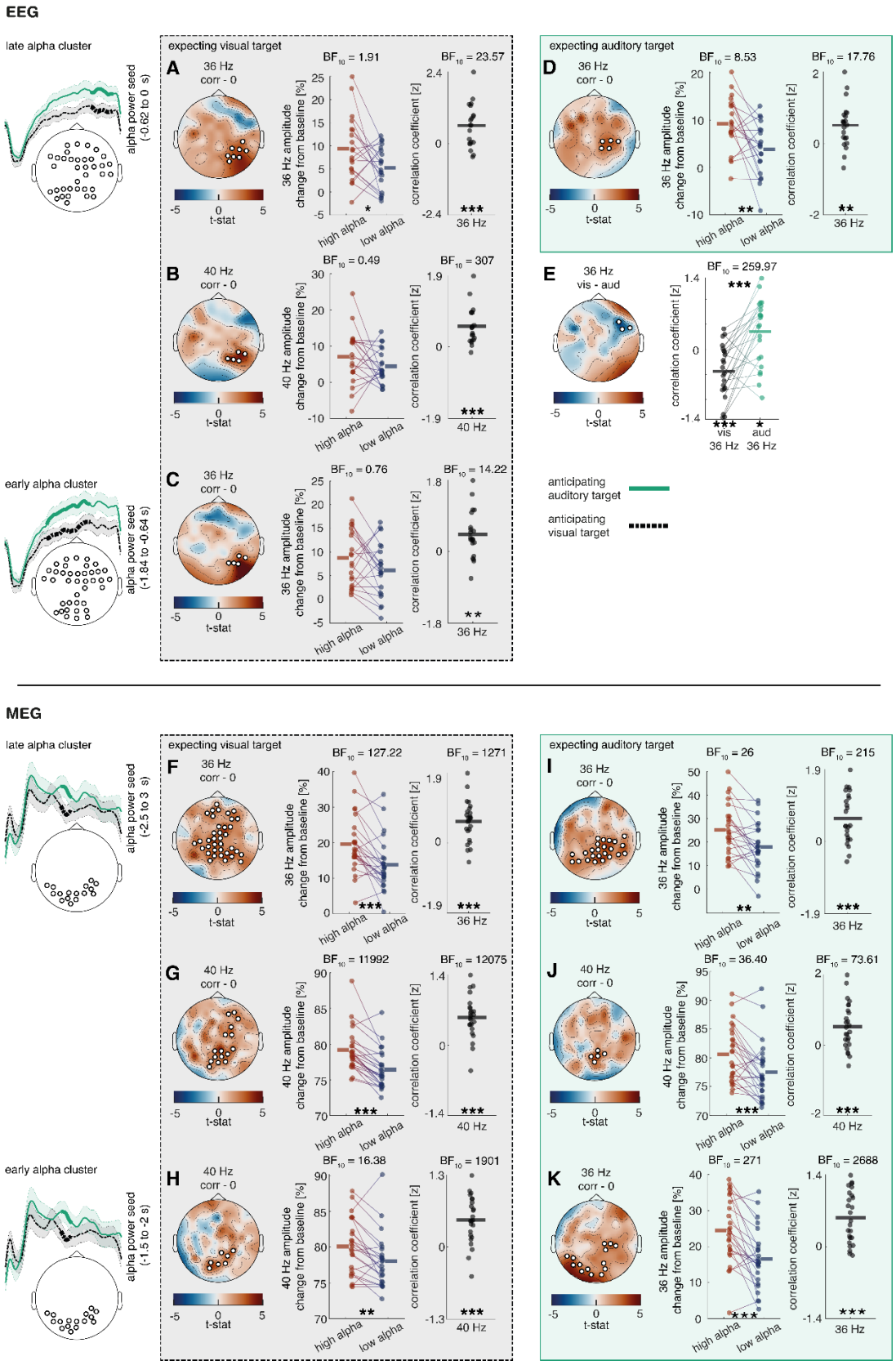


Figure 5. Relationship between cue induced alpha modulation and amplitude of frequency tagged responses. Previously obtained alpha clusters (see Fig. 3) were correlated over trials with 40 Hz and 36 Hz clusters (see Fig. 4), where alpha electrodes/sensors were applied as seeds. The analysis was performed using a cluster-permutation approach, testing a correlation model against a 0-correlation model. Clusters significantly diverging from the 0-correlation model are presented topographically.

Additionally, median splits between high and low alpha trials as well as correlation coefficients of these clusters are displayed for all participants **A**-**B,** a positive correlation is visible between alpha activity in the last 400 ms and steady state potentials shortly before target onset when expecting a visual target (36 HZ: p = .013; 40 Hz: p = .009). **D**, when expecting an auditory target, there is a positive correlation visible between alpha activity in the last 400 ms and 36 Hz activity shortly before target onset (p = .010). **E,** the correlation between alpha activity 400 ms and 36 Hz activity shortly before target onset changes its direction depending on whether an auditory or a visual target is expected (p = .037). **C**, a positive correlation is also visible between alpha activity as early as ~1200 ms to 400 ms and 36 Hz activity shortly before target onset when expecting a visual target (p = .016). **F-H**, a positive correlation is visible between alpha activity in the last 500 ms as well as alpha activity in the last 1500ms–1000 ms and steady state potentials shortly before target onset when expecting a visual target (36 HZ late: p = .013; 40 Hz late: p = .009; 40 Hz early: p = 002). **I-K**, when expecting an auditory target, there is a positive correlation between alpha activity in the last 500 ms as well as alpha activity in the last 1500ms–1000 ms and steady state potentials shortly before target onset (36 HZ late: p < .001; 40 Hz late: p = .005; 36Hz early: p = 011). **A-K**, EEG: N = 22; MEG: N = 27; *** sig < .001; ** sig. < .01; * sig. < .05. + sig. < .1

Intermodulation frequency

Lastly, we analysed the steady-state response of the intermodulation frequency at 4 Hz. Increased intermodulation shortly before target onset could be observed when expecting an auditory compared to a visual target (-0.51 to -0.0620 s; p < .001). This effect was strongest over left fronto-to-central electrodes and right central-to-occipital electrodes (see Fig. 6A). The same increase in the intermodulatory frequency could be observed in our second study, during the last 500 ms prior to target onset (see Fig. 6C; p = .006). In source space, a descriptive condition difference was visible in auditory sensory cortices, however this effect did not reach significance (p = .49; peak spm coordinates: 45 -83 -19 mm, in the right lingual gyrus; see Fig. 6D). With the goal to examine whether there are any behaviourally effects of the condition difference, trial by trial correlations with reaction times of each of the 2 auditory and visual conditions were performed. Only in the easiest condition, when expecting a visual target that was not accompanied by a distractor, a negative correlation with reaction time could be found, strongest over right central electrodes (see Fig. 6B; p = .046). While a median split between slow and fast trials did not reach significance (p = .50; $t_{(21)}$ = -0.69; $BF_{(10)}$ = .28), testing the

correlation coefficients against 0 revealed strong evidence for a correlation (p = .004; $t_{(21)}$ = -3.28; $BF_{(10)}$ = 11.98).

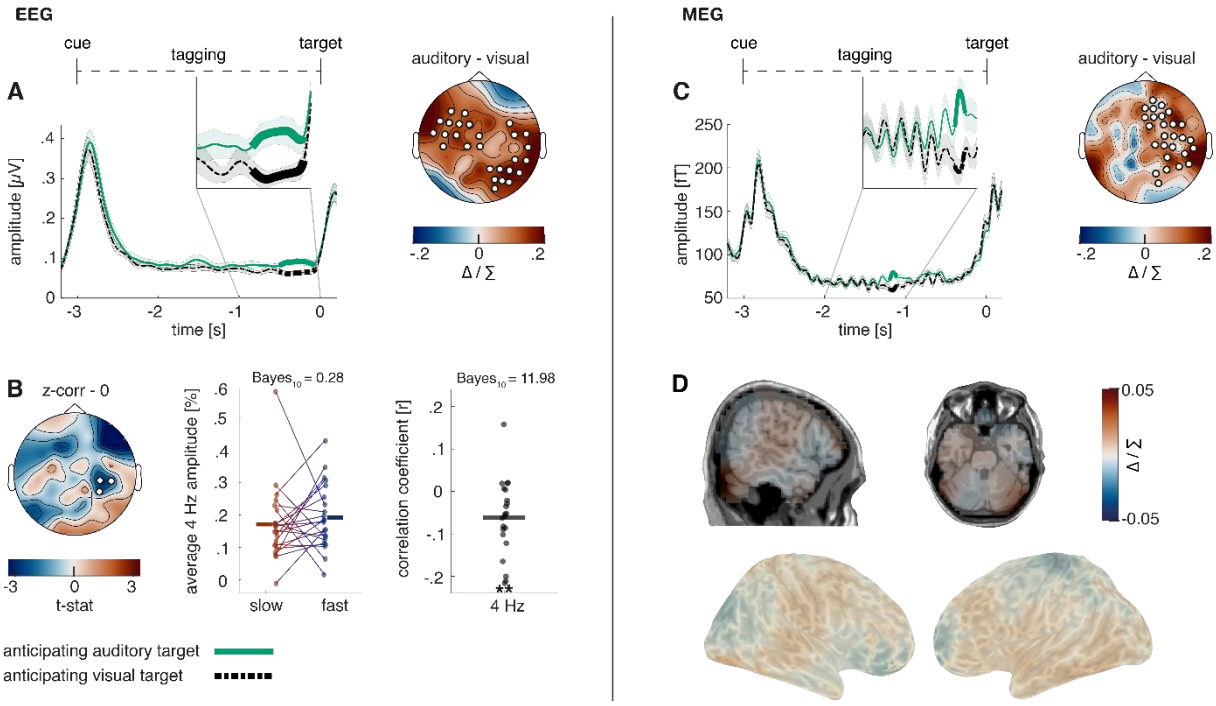


Figure 6. Steady-state response in the intermodulation frequency and its behavioural relevance. A, the Hilbert-envelope of the 4 Hz steady-state response reveals an increase shortly before target onset when anticipating an auditory compared to a visual target (p < .01). **B,** there is a trial-by-trial correlation between 4 Hz activity and reaction time when a visual target without distractor was presented. The correlation is further illustrated by a median split between fast and slow reaction time trials as well as by correlation coefficients for each participant. **C**, replication of the results presented in (A) in our MEG-study (p = .006). **D**, source localization showed activity over auditory sensory areas, but did not reach significance. EEG: N = 22; MEG: N = 27; ** sig. < .01;

Discussion

The neuropsychological account of attention defines it as the selective facilitation (i.e., prioritization) of relevant sensory input and suppression of irrelevant sensory input. Oscillatory activity in the alpha range (~10 Hz) has been suggested to play a mechanistic role in attention through inhibition of irrelevant cortices, commonly referred to as the 'alpha inhibition hypothesis' (Foxe et al., 1998; Jensen & Mazaheri, 2010; Klimesch et al., 2007). In the current cross-modal attention study we directly tested this hypothesis by using frequency-tagging to specifically examine how cues signalling the modality of an upcoming target (either the auditory or visual modality) affected the responsiveness of the relevant and irrelevant sensory cortices prior to target onset. In-line with previous work, we observed a post-cue increase in posterior alpha power in anticipation of processing auditory targets. However, contrary to prevalent theories proposing visual gain suppression when focusing on the auditory modality, we observed that the amplitude of visual frequency-tagging responses increased just prior to the onset of the auditory target. This suggests that responsiveness of the visual stream was not inhibited when attention was directed to auditory processing and was not inhibited by occipital alpha activity. Our results reconcile previously paradoxical results on audio-visual attention and support the view that alpha activity gates downstream communication pathways.

Frequency-tagging

In the current experiment, we specifically chose to analyse the cue-to-target interval, where both visual and auditory SSEPs/SSEFs present preparatory states for the upcoming task, independent of task-related processing or performance. The magnitude of auditory SSEPs/SSVEFs was increased shortly before target onset when expecting a demanding auditory target compared to a visual target, very much in line with previous reports (e.g., Saupe et al., 2009). In contrast to the results reported in Saupe et a., (2009), where visual SSEPs decreased when attending the auditory modality, visual SSEPs/SSVEFs increased shortly before target

onset when expecting an *auditory* target in our data. This is especially surprising as auditory targets were frequently or in case of our second study, always accompanied by visual distractors, rendering it optimal for task success to completely ignore any visual input. As auditory targets were significantly more difficult than visual targets in our first study and of comparable difficulty in our second study, these results strongly speak to a vigilance increase of sensory processing independent of modality and an inability to selectively disengage one sensory modality in anticipation of a demanding task. This view is consistent with previous work in which visual SSEPs elicited by irrelevant background stimulation increased with task load in an auditory discrimination task (Jacoby et al., 2012). Furthermore, our results indicate that task demand is a strong candidate to reconcile previously seemingly paradox results, as splitting attention between the auditory and visual system seemed possible in simpler tasks (Driver & Spence, 1998; Saupe et al., 2009) and impossible under high demand (de Jong et al., 2010; Driver, 1996; Driver & Spence, 1998; Spence & Driver, 1996). An alternative account for our findings stems from the evidence, that participants are more likely to only perceive and react to the visual modality, when confronted with audio-visual stimuli (Colavita, 1974; Spence, 2009). However, this effect was mostly limited to speeded modality discrimination/target detection tasks (Sinnett et al., 2008; Spence, 2009). Furthermore, the increased difficulty of the here-used auditory stimuli was confirmed in a previous block-design study (van Diepen & Mazaheri, 2017) and in our second study, performances over the visual and auditory tasks were comparable. Nevertheless, visual dominance could play a role for auditory target difficulty as well as predictions over the reciprocity of the audio-visual relationship.

A revision of the Alpha inhibition hypothesis

Top-down cued changes in alpha power have now been widely viewed to play a functional role in directing attention: the processing of irrelevant information is attenuated by increasing alpha power in cortices involved with processing this information (Foxe, Simpson, & Ahlfors, 1998;

Hanslmayr et al., 2007; Jensen & Mazaheri, 2010). However, recent evidence suggests that alpha activity does not inhibit gain in early sensory processing stages (Antonov et al., 2020; Gundlach et al., 2020; Gutteling et al., 2022; Zhigalov & Jensen, 2020). To date there has been no direct investigation into the effect of alpha increases on later stages in the processing stream. In the current study, as expected, we observed a post-cue increase in occipital alpha activity in anticipation of an auditory target. However, we also observed an increase in the amplitude of visual SSEPs during the cue-target interval, which directly contradicts the widespread view of alpha activity exerting 'gain control' in early sensory areas by regulating excitability (Foxe & Snyder, 2011; Jensen & Mazaheri, 2010; Van Diepen et al., 2019). Here we propose that alpha activity, rather than modulating early primary sensory processing, exhibits its inhibitory effects at later stages of the processing stream (Antonov et al., 2020; Gundlach et al., 2020; Zhigalov & Jensen, 2020; Zumer et al., 2014), gating feedforward or feedback communication between sensory areas (Bauer et al., 2020; Haegens et al., 2015; Uemura et al., 2021). Our data provides evidence in favour of this view, as we can show that alpha activity covaries over trials with SSEP magnitude in adjacent areas. If alpha activity exerted gain control in early visual regions, increased alpha activity would have to lead to a decrease in SSEP responses. In contrast, we observe that increased alpha activity originating from early visual cortex is related to *enhanced* visual processing at later stages of the processing stream, which we could confirm using source analysis. It seems plausible to assume that inhibition of other task-irrelevant communication pathways leads to prioritised and thereby enhanced processing over relevant pathways. In line with previous literature, we therefore suggest that alpha activity limits task-irrelevant feedforward communication, thereby enhancing processing capabilities in relevant downstream areas (see Fig. 1A). Furthermore, we could show that the magnitude of the correlation between alpha power and visual information processing varied between conditions, suggesting a dynamic and adaptive network.

It is known that the localisation of alpha activity reflects the retinotopic organisation of visual spatial attention topographically over the parietooccipital cortex (Kelly et al., 2006; Popov et al., 2019). Notably, recent studies provided evidence, that the same organisation can be observed for auditory attention. Specifically, the localisation of visual alpha activity in the parietooccipital cortex reflects the spatial direction of auditory attention (Klatt et al., 2021; Popov et al., 2021). This observation can be explained through micro-saccades towards the spatial location of sounds, which are irrevocably connected to alpha oscillations (Popov et al., 2021). While we did not manipulate spatial attention, our results fit well to the notion of visual alpha activity serving as a sensory orientation system, relaying visual information to task-relevant downstream processing areas, and blocking communication to irrelevant pathways.

The intermodulation frequency

Previous research showed that simultaneous frequency-tagging in multiple frequencies evokes a response in the intermodulation frequency (f1 – f2). In multimodal settings, this frequency is thought to reflect cross-modal integration (Drijvers et al., 2021). This is very well in line with our findings, where increased vigilance of the sensory system arising from anticipation of a difficult auditory target resulted in an increase in the intermodulatory frequency. Furthermore, we could show that this frequency covaries over trials with reaction time in the easiest condition, where visual targets were presented without any distractors. A lack of this connection in other conditions might reflect increasing interferences from higher task difficulty, rather than a lack of the effect itself, but this remains to be tested. We cannot exclude an alternative explanation, as theta oscillations are known to be involved in movement preparation, it is possible that phase-resets could lead to time-locked appearance of these oscillations (Lakatos et al., 2008; Tomassini et al., 2017).

Conclusion

Our results taken together suggest that under high task difficulty, audio-visual excitability is enhanced, reflecting an increase in vigilance for the sensory system, even if this increases processing of distracting information. We showed that this vigilance shift, as reflected by SSEP/SSEF responses, is regulated by alpha activity, presumably through relaying of sensory information over communication pathways, thereby controlling the downstream flow of sensory information.

Materials and Methods

Participants EEG-Study

In total, 24 healthy volunteers participated in this study (mean age: 19.1 ± 1.8 SD; 17 women). Due to technical difficulties, one participant could not finish the experiment and one participant did not exceed chance level in the behavioural task (~33 %). Both were therefore removed from any further analysis. All remaining participants reported normal or corrected-to-normal vision, no history of psychiatric or neurological illness and provided written informed consent. After completion of the experiment, participants received either monetary compensation or certification of their participation for their university course program. The study protocol was approved by the Ethics Committee of the School of Psychology at the University of Birmingham and is in accordance with the Declaration of Helsinki.

Participants MEG-Study

In total, 28 healthy volunteers participated in this study (mean age: 23.4 ± 3.6 SD; 20 women). One participant was removed from further analysis, as they only responded to ~42% of trials correctly in the second block, which related to 27/19 trials per condition respectively. All remaining participants reported normal or corrected-to-normal vision, no history of psychiatric or neurological illness and provided written informed consent. After completion of the experiment, participants received either monetary compensation or certification of their participation for their university course program. The study protocol was approved by the Ethics Committee of the School of Psychology at the University of Birmingham and is in accordance with the Declaration of Helsinki.

Cross-modal attention paradigm EEG-Study

Each trial was initiated by a brief presentation of a cue (100 ms) signalling the modality of the upcoming discrimination task (v-shape: visual modality; inversed v-shape: auditory modality; diamond-shape: unspecified). During the following cue-to-target interval, the fixation cross was frequency-tagged at 36 Hz. At the same time, a 40 Hz frequency-tagged sound (amplitude modulated white noise) was played over headphones (see Fig. 1). The volume of tones was initially adjusted to a level that was clearly perceivable but not uncomfortable and remained stable over participants. No task was connected to this interval and participants did not need to

pay attention to either the sound or the fixation cross. After three seconds, the frequency-tagging stopped, and right after the cessation of stimuli, the target was presented for a very brief moment (25 ms). It consisted either of a Gabor patch (visual modality) or a sound (auditory modality). If the target modality was visual, participants had to use the three arrow buttons on the keyboard to indicate whether the Gabor patch was tilted to the left (-10°; left arrow button), vertical (0°; down arrow button), or tilted to the right (10°; right arrow button). Additionally, in 50% of the trials, a random distractor from the pool of auditory targets was presented simultaneously to the visual target over headphones. Similarly, if the target modality was auditory, participants used the same buttons to indicate whether the pitch of the tone was low (500 Hz), medium (1000 Hz) or high (2000 Hz). Again, in 50% of the trials, a random distractor from the pool of visual targets was also presented. If the cue was unspecified (diamond shape), either a visual (50% of unspecified trials) or an auditory target was presented, never accompanied by any distractors. Experimental trials were separated by an inter-trial interval of 4 seconds, to avoid carry-over effects from previous trials. The resulting 6 conditions were randomly ordered and balanced out over the experiment.

During the experiment, participants were instructed to keep their gaze locked to a fixation cross presented at the centre of the screen. Preceding data collection, participants performed 36 practice trials to get accustomed to the task and the target stimuli. The ensuing experiment was split into 26 trial sequences, separated by self-chosen breaks, which together resulted in 468 trials and lasted between 80 and 90 minutes. The discrimination task was programmed and presented with MATLAB® R2020b and Psychtoolbox-3 on an LCD-monitor featuring a 140 Hz refresh rate. The onset of the visual and auditory tagging frequencies (i.e steady state stimuli) were tracked using the Cedrus Stimtracker (https://cedrus.com/stimtracker/index.htm).

Adjustments to the attention paradigm in the MEG-Study

In our second study, we removed all ambiguity concerning targets and distractors and therefore developed a blocked design, incorporating two blocks. The first block did not display distractors and only correctly predicting target cues were presented. Cues in the second block were likewise always correctly indicating the target modality, but this time, each target was accompanied by a random distractor from the non-target modality.

Furthermore, the visual task was adjusted to be more in line with the auditory task. As such, the Gabor patches now featured stripes in different frequencies (e.g. a low number of stripes, a

medium number of stripes and a high number of stripes. The participant's task was to discriminate between these three Gabor patches. As auditory targets had been markedly more difficult in our first study, we now included a brief difficulty calibration prior to the experiment. First, we presented 21 Gabor patches with 3 different amounts of stripes following a standard difficulty. If participants could discriminate them correctly 55 – 75% of the time, this difficulty setting was chosen. Otherwise, depending on the performance, the stripe-frequency of the Gabor patches was adjusted. There were maximally 3 sessions of 21 Gabor patches, after which we had enough data to calibrate the individual difficulty setting.

The same procedure was then performed with the difficulty of the tones, calibrating the pitch frequency for each individual participant.

Lastly, visual frequency-tagging stimulation now followed a sinusoidal contrast-change rather than an on-off stimulation, which was possible due to a high-resolution projector featuring a refresh rate of 1440 Hz (PROPixx DLP LED projector ;VPixx Technologies Inc., Canada).

Eye-tracking

To make sure participants focused on the fixation cross during the cue-to-target interval, we incorporated eye-tracking into our MEG-experiment (EyeLink 1000 Plus). Correct trials of the second block were analysed for vertical and horizontal eye-movements. To remove blinks, trials with very large eye-movements (> 10 degrees of visual angle) were removed from the data (See suppl Fig. 5).

Behavioral analysis

We were interested in accuracy in the discrimination of visual targets and auditory targets, as well as reaction times. Furthermore, we examined the distraction cost of having a target presented with a distractor of a different modality as well as the reaction time to make the target discrimination. The distraction cost was calculated as the reaction time difference between cued targets with distractors (i.e. visual and auditory stimuli presented together) and cued targets without distractors (either a visual or auditory stimulus presented alone). All incorrect trials as well as trials with reaction times faster than 100 ms or exceeding 1500 ms were removed from analysis (0.5 % too fast, 9.2% too slow).

EEG data acquisition

All EEG recordings were conducted using a WaveGuard Cap (ANTneuro), featuring 64 Ag/AgCL electrodes (10-10 system; ground: Fz; reference: Cpz; EOG: left canthus). Electrodes positions were prepared with OneStep cleargel conductive paste and impedances were kept below 100 kΩ. The measured signal was transmitted using an ANTneuro EEGosports amplifier (low-pass filter: 150 Hz; high-pass filter: 0.5 Hz; sampling rate: 500 Hz).

MEG data acquisition

Prior to the experiment, feducial positions and head-shape were recorded using a FASTRAK system (Polhemus, USA). The experiment took place in a dimly lit room, where participants were seated in a comfortable chair in the gantry of a 306-sensor TRIUX Elekta system with 204 orthogonal planar gradiometers and 102 magnetometers (Elekta, Finland). The 71*40 cm screen was positioned at ~1.40 m distance from the participant.

EEG Preprocessing

Offline analyses were performed in MATLAB ® R2020b. The data was pruned from artifacts by visual inspection using the EEGLAB toolbox (Delorme & Makeig, 2004). Additionally, blinks and ocular artefacts were removed from the data using independent component analysis (ICA). EEG channels were re-referenced to an average of all channels (excluding EOG).

MEG Preprocessing

Offline analyses were performed in MATLAB ® R2020b and Python. Spatiotemporal Signal-Source-Separation (SSS) was applied to the raw data via MNE's inbuilt maxfilter function with a duration window of 10 s and a correlation value of .9. The data was pruned from artifacts by visual inspection using the Fieldtrip toolbox (Delorme & Makeig, 2004). Additionally, blinks and ocular artefacts were removed from the data using independent component analysis (ICA). In sensor space, planar gradiometers were combined for further analyses. In source space, all individual planar gradiometers were analysed.

Amplitude of the evoked frequency-tagging response

To investigate the temporal dynamics of amplitude of the frequency tagged responses after the onset of the attentional cues (also the precise onset of the frequency tagged stimuli) the data was epoched into 6-second segments starting 1.5 seconds prior to cue onset. Next the data were narrow-band filtered around the 36 Hz activity to capture the visual frequency-tagging, 40 Hz activity in to capture the auditory frequency-tagging, and the intermodulation frequency at 4

Hz, which can be derived by subtracting both frequency-tagging responses ($f_i = f_{auditory} - f_{visual}$; see Drijvers, Spaak & Jensen, 2020). Here we used a Blackmann-windowed sync filters adapted to a suitable ratio of temporal and frequency resolution for the specific frequency of each of the tagged signals: filter order 116 for 35.5 to 36.5 Hz and 39.5 to 40.5, filter order 344 for 3.5 to 4.5 Hz. The filtered data at each of the tagged frequencies as well the intermodulated frequency were baseline corrected (interval between 700 and 200 ms preceding cue onset) before calculating the average over trials to obtain steady-state evoked potentials). The power envelope of the SSEPs of tagged frequencies was estimated using Hilbert transformation.

Temporal dynamics of the induced EEG changes

In addition to looking at the cue evoked changes in the amplitude of the frequency tagged signals, we investigated the induced changes in the EEG signal at the frequencies of the tagged auditory and visual stimuli, alpha activity (9-11 Hz, filter order 276 for 7.5 to 12.5 Hz), as well as the intermodulation frequency (4 Hz; filter order 344 for 3.5 to 4.5 Hz). Here rather than averaging the epoched data filtered at the specific frequency ranges, we performed the Hilbert transform, and averaged the power-envelope of the specific frequencies across trials. This approach is very much analogous to the standard time-frequency analysis using convolutions (van Diepen & Mazaheri, 2017; Zhigalov et al., 2019), but affords more control concerning temporal versus frequency resolution to examine the temporal dynamics of the specific frequencies of interest. In our second study, the individual peak alpha frequency was used for bandpass-filtering in contrast to a standardised band applied in the first study.

Time-frequency representations of power

In addition to the estimating frequency power envelopes, Time–frequency representations (TFRs) of power of the EEG signal were estimated using the Fieldtrip toolbox (Oostenveld et al., 2011). The power or frequencies between 5 and 20 Hz were calculated for each trial, using a sliding time window (frequency steps: 0.5; time steps: 10 ms). The length of the window was adjusted to a length of 3 cycles per frequency and tapered with a Hanning window. For each trial, both datasets were normalised to display relative percent change from baseline using the following formula: [(activity – baseline) / baseline], where baseline refers to the interval between 700 and 200 ms before cue onset. To estimate the topographical distribution of voltage differences between conditions, uncorrected power values were normalised applying the following formula: [$\Delta/\sum = (a - b) / (a + b)$], where a and b reflect the different conditions.

*Source localization*Source localization was performed with a beamformer approach using the Fieldtrip toolbox. Headmodels were created based on individual T1-scans fitted to fiducial points and head shapes. These data were fit to a 5mm 3d sourcemodel and warped into MNI-space. Two participants were missing individual T1-scans. In these cases, we applied a standardized T1-scan using the Colin 27 Average Brain Model (Holmes et al., 1998). Frequency-domain data was localized using the Dynamic Imaging of Coherent Sources (DICS beamformer) method with a dpss taper of 2 Hz assuming fixed orientation. As condition differences between frequency-tagging responses were better estimated in the time-domain, we assessed them applying the Synthetic Aperture Magnetometry (SAM-beamformer) method with optimal fixed rotation (Sekihara et al., 2004). Significant brain areas and peak coordinates were related to brain areas using the Anatomical Automatic Labeling (AAL) atlas for SPM8 (Tzourio-Mazoyer et al., 2002) . After statistical analysis, source localized data was interpolated onto the Colin 27 Average Brain Model MRI. Cerebellar and brainstem interpolations were excluded from the coordinate system.

Statistical analysis

Condition differences in the behavioural task were estimated with paired t-tests and repeated measures ANOVAS, utilizing Tukey-Kramer Post-Hoc test. Power differences between conditions, as well as source-space contrasts were analysed using cluster permutation analysis (Maris & Oostenveld, 2007). In this procedure, condition labels were randomly shuffled 1000 times, creating pairs of surrogate conditions. To test for significance, paired t-tests were conducted for each data point and each channel, resulting in one t- matrix for real conditions and 1000 t-matrices for surrogate conditions. Significant t-values ($p < .05$) were defined as clusters if there was at least one significant data point present at the same time and frequency in at least two neighbouring channels. To correct for multiple comparisons, a condition difference was only assumed, if the maximum sum of t-values in a real cluster exceeded the same sum of 95% of the clusters found in the surrogate data. To replicate our results in the second study, we applied the same statistics and averaged the previously found time-intervals into windows of 500ms (e.g.: if we found an effect -0.51 to -0.0620 s prior to target onset we tested this effect for a time window of -0.5 to 0 s prior to target onset).

The relationship between induced changes in alpha activity and frequency tagged responses was assessed using trial by trial Spearman correlations. For each participant and each electrode, a correlation coefficient was calculated between the average activity in a previously identified

cluster, which was used as seed (e.g., condition differences in alpha activity), and the average activity of the electrophysiological correlate of interest (e.g., 36 Hz activity over the previously identified time window). The correlation coefficients were z-transformed, and the resulting channel by participant matrix was tested against null-correlation model using the cluster permutation approach described above. Derived clusters were additionally tested and visualised by comparing median split trials of high vs low activity. For this analysis, outliers (values deviating more than 2 standard deviations from the mean) were excluded. Furthermore, the average correlation coefficient of the cluster was tested against a 0-correlation model for each participant using t-statistics. Lastly, an interaction between electrophysiological correlations and conditions was performed using correlation coefficients for each participant and electrode, testing them between conditions using the cluster permutation approach. Perceptually uniform and universally readable colormaps were applied to all visualisations (Crameri et al., 2020). All data are presented as mean ± standard error of the mean (SEM).

Acknowledgements: This work was made possible by funding support from Facebook Oculus and BBSRC (BB/R018723/1).

Competing interests: There are no competing interests

Data availability: Codes for analyses and figures will be made available on github and the data will be uploaded to Dryad or OSF

References

Antonov, P. A., Chakravarthi, R., & Andersen, S. K. (2020). Too little, too late, and in the wrong place: Alpha band activity does not reflect an active mechanism of selective attention. *NeuroImage*, *219*, 117006. https://doi.org/10.1016/J.NEUROIMAGE.2020.117006

Bauer, A. K. R., Debener, S., & Nobre, A. C. (2020). Synchronisation of Neural Oscillations and Cross-modal Influences. In *Trends in Cognitive Sciences* (Vol. 24, Issue 6, pp. 481–495). Elsevier Ltd. https://doi.org/10.1016/j.tics.2020.03.003

Brickwedde, M., Bezsudnova, Y., Kowalczyk, A., Jensen, O., & Zhigalov, A. (2022). Application of rapid invisible frequency tagging for brain computer interfaces. *Journal of Neuroscience Methods*, *382*, 109726. https://doi.org/10.1016/J.JNEUMETH.2022.109726

Brickwedde, M., Schmidt, M. D., Krüger, M. C., & Dinse, H. R. (2020). 20 Hz Steady-State Response in Somatosensory Cortex During Induction of Tactile Perceptual Learning Through LTP-Like Sensory Stimulation. *Frontiers in Human Neuroscience*. https://doi.org/10.3389/fnhum.2020.00257

Colavita, F. B. (1974). Human sensory dominance*. In *Perception & Psychophysics* (Vol. 16, Issue 2).

Colon, E., Legrain, V., & Mouraux, A. (2012). Steady-state evoked potentials to study the processing of tactile and nociceptive somatosensory input in the human brain. *Neurophysiologie Clinique/Clinical Neurophysiology*, *42*(5), 315–323. https://doi.org/10.1016/J.NEUCLI.2012.05.005

Crameri, F., Shephard, G. E., & Heron, P. J. (2020). The misuse of colour in science communication. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-19160-7

de Jong, R., Toffanin, P., & Harbers, M. (2010). Dynamic crossmodal links revealed by steady-state responses in auditory–visual divided attention. *International Journal of Psychophysiology*, *75*(1), 3–15. https://doi.org/10.1016/J.IJPSYCHO.2009.09.013

Dinse, H. R., Ragert, P., Pleger, B., Schwenkreis, P., & Tegenthoff, M. (1998). Pharmacological Modulation of Perceptual Learning and Associated Cortical Reorganization. *SCIENCE*, *282*, 865.

Drijvers, L., Jensen, O., & Spaak, E. (2021). Rapid invisible frequency tagging reveals nonlinear integration of auditory and visual information. *Human Brain Mapping*, *42*(4), 1138–1152. https://doi.org/10.1002/hbm.25282

Driver, J. (1996). Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading. *NATURE ·*, *381*.

Driver, J., & Spence, C. (1998). Cross-modal links in spatial attention. *Philos Trans R Soc Lond B Biol Sci*, *353*, 1319–1331.

Dugué, L., Marque, P., & VanRullen, R. (2011). The phase of ongoing oscillations mediates the causal relation between brain excitation and visual perception. *Journal of Neuroscience*, *31*(33), 11889–11893. https://doi.org/10.1523/JNEUROSCI.1161-11.2011

Ergenoglu, T., Demiralp, T., Bayraktaroglu, Z., Ergen, M., Beydagi, H., & Uresin, Y. (2004). Alpha rhythm of the EEG modulates visual detection performance in humans. *Cognitive Brain Research*, *20*(3), 376–383. https://doi.org/10.1016/J.COGBRAINRES.2004.03.009

Foxe, J. J., Simpson, G. V, & Ahlfors, S. P. (1998). Parieto-occipital ~10Hz activity reflects anticipatory state of visual attention mechanisms. *NeuroReport*.

Foxe, J. J., & Snyder, A. C. (2011). The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. *Frontiers in Psychology*, *2*(JUL). https://doi.org/10.3389/fpsyg.2011.00154

Fu, K. M. G., Foxe, J. J., Murray, M. M., Higgins, B. A., Javitt, D. C., & Schroeder, C. E. (2001). Attention-dependent suppression of distracter visual input can be cross-modally cued as indexed by anticipatory parieto–occipital alpha-band oscillations. *Cognitive Brain Research*, *12*(1), 145–152. https://doi.org/10.1016/S0926-6410(01)00034-9

Gundlach, C., Moratti, S., Forschack, N., & Müller, M. M. (2020). Spatial Attentional Selection Modulates Early Visual Stimulus Processing Independently of Visual Alpha Modulations. *Cerebral Cortex*, *30*(6), 3686–3703. https://doi.org/10.1093/cercor/bhz335

Gutteling, T. P., Sillekens, L., Lavie, N., & Jensen, O. (2022). Alpha oscillations reflect suppression of distractors with increased perceptual load. *Progress in Neurobiology*, *214*. https://doi.org/10.1016/J.PNEUROBIO.2022.102285

Haegens, S., Barczak, A., Musacchia, G., Lipton, M. L., Mehta, A. D., Lakatos, P., & Schroeder, C. E. (2015). Laminar profile and physiology of the α rhythm in primary visual, auditory, and somatosensory regions of neocortex. *Journal of Neuroscience*, *35*(42), 14341–14352. https://doi.org/10.1523/JNEUROSCI.0600-15.2015

Haegens, S., Nácher, V., Luna, R., Romo, R., & Jensen, O. (2011). α-Oscillations in the monkey sensorimotor network influence discrimination performance by rhythmical inhibition of neuronal spiking. *Proceedings of the National Academy of Sciences*, *108*(48), 19377–19382. https://doi.org/10.1073/PNAS.1117190108

Hari, R., Hämäläinen, M., & Joutsiniemi, S. L. (1989). Neuromagnetic steady-state responses to auditory stimuli. *Journal of the Acoustical Society of America*, *86*(3), 1033–1039. https://doi.org/10.1121/1.398093

Holmes, C. J., Hoge, R., Collins, L., Woods, R., Toga, A. W., & Evans, A. C. (1998). Enhancement of MR images using registration for signal averaging. *Journal of Computer Assisted Tomography*, *22*(2), 324–333. https://doi.org/10.1097/00004728-199803000-00032

Jacoby, O., Hall, S. E., & Mattingley, J. B. (2012). A crossmodal crossover: Opposite effects of visual and auditory perceptual load on steady-state evoked potentials to irrelevant visual stimuli. *NeuroImage*, *61*(4), 1050–1058. https://doi.org/10.1016/J.NEUROIMAGE.2012.03.040

Jensen, O., & Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: Gating by inhibition. *Frontiers in Human Neuroscience*, *4*. https://doi.org/10.3389/fnhum.2010.00186

Kelly, S. P., Lalor, E. C., Reilly, R. B., & Foxe, J. J. (2006). Increases in alpha oscillatory power reflect an active retinotopic mechanism for distracter suppression during sustained visuospatial attention. *Journal of Neurophysiology*, *95*(6), 3844–3851. https://doi.org/10.1152/jn.01234.2005

Klatt, L.-I., Getzmann, S., & Schneider, D. (2021). Attentional Modulations of Alpha Power Are Sensitive to the Task-relevance of Auditory Spatial Information. *BioRxiv*. https://doi.org/10.1101/2021.02.12.430942

Klimesch, W., Sauseng, P., & Hanslmayr, S. (2007). EEG alpha oscillations: The inhibition–timing hypothesis. *Brain Research Reviews*, *53*(1), 63–88. https://doi.org/10.1016/J.BRAINRESREV.2006.06.003

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science (New York, N.Y.)*, *320*(5872), 110–113. https://doi.org/10.1126/SCIENCE.1154735

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/J.JNEUMETH.2007.03.024

Marzoll, A., Saygi, T., & Dinse, H. R. (2018). The effect of LTP- and LTD-like visual stimulation on modulation of human orientation discrimination. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-34276-z

Mazaheri, A., van Schouwenburg, M. R., Dimitrijevic, A., Denys, D., Cools, R., & Jensen, O. (2014). Region-specific modulations in oscillatory alpha activity serve to facilitate processing in the visual and auditory modalities. *NeuroImage*, *87*, 356–362. https://doi.org/10.1016/J.NEUROIMAGE.2013.10.052

Müller, M. M., & Hillyard, S. (2000). Concurrent recording of steady-state and transient event-related potentials as indices of visual-spatial selective attention. *Clinical Neurophysiology*, *111*(9), 1544–1552. https://doi.org/10.1016/S1388-2457(00)00371-0

Müller, M. M., Picton, T. W., Valdes-Sosa, P., Riera, J., Teder-Sälejärvi, W. A., & Hillyard, S. A. (1998). Effects of spatial selective attention on the steady-state visual evoked potential in the 20–28 Hz range. *Cognitive Brain Research*, *6*(4), 249–261. https://doi.org/10.1016/S0926-6410(97)00036-0

Okazaki, Y. O., De Weerd, P., Haegens, S., & Jensen, O. (2014). Hemispheric lateralization of posterior alpha reduces distracter interference during face matching. *Brain Research*, *1590*(1), 56–64. https://doi.org/10.1016/J.BRAINRES.2014.09.058

Pantev, C., Roberts, L. E., Elbert, T., Roß, B., & Wienbruch, C. (1996). Tonotopic organization of the sources of human auditory steady-state responses. *Hearing Research*, *101*(1–2), 62–74. https://doi.org/10.1016/S0378-5955(96)00133-5

Popov, T., Gips, B., Kastner, S., & Jensen, O. (2019). Spatial specificity of alpha oscillations in the human visual system. *Human Brain Mapping*, *40*(15), 4432–4440. https://doi.org/10.1002/hbm.24712

Popov, T., Gips, B., Weisz, N., & Jensen, O. (2021). Brain areas associated with visual spatial attention display topographic organization during auditory spatial attention. *BioRxiv*. https://doi.org/10.1101/2021.03.15.435371

Porcu, E., Keitel, C., & Müller, M. M. (2013). Concurrent visual and tactile steady-state evoked potentials index allocation of inter-modal attention: A frequency-tagging study. *Neuroscience Letters*, *556*, 113–117. https://doi.org/10.1016/J.NEULET.2013.09.068

Regan, D. (1982). COMPARISON OF TRANSIENT AND STEADY-STATE METHODS*. *Annals of the New York Academy of Sciences*.

Saupe, K., Schröger, E., Andersen, S. K., & Müller, M. M. (2009). Neural mechanisms of intermodal sustained selective attention with concurrently presented auditory and visual stimuli. *Frontiers in Human Neuroscience*, *3*(NOV). https://doi.org/10.3389/neuro.09.058.2009

Sekihara, K., Nagarajan, S. S., Poeppel, D., & Marantz, A. (2004). Asymptotic SNR of scalar and vector minimum-variance beanformers for neuromagnetic source reconstruction. *IEEE Transactions on Biomedical Engineering*, *51*(10), 1726–1734. https://doi.org/10.1109/TBME.2004.827926

Sinnett, S., Soto-Faraco, S., & Spence, C. (2008). The co-occurrence of multisensory competition and facilitation. *Acta Psychologica*, *128*(1), 153–161. https://doi.org/10.1016/J.ACTPSY.2007.12.002

Snyder, A. (1992). Steady-state vibration evoked potentials: description of technique and characterization of responses. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *84*(3), 257–268. https://doi.org/10.1016/0168-5597(92)90007-X

Spaak, E., Bonnefond, M., Maier, A., Leopold, D. A., & Jensen, O. (2012). Layer-Specific Entrainment of Gamma-Band Neural Activity by the Alpha Rhythm in Monkey Visual Cortex. *Current Biology*, *22*(24), 2313–2318. https://doi.org/10.1016/J.CUB.2012.10.020

Spence, C. (2009). Explaining the Colavita visual dominance effect. In *Progress in Brain Research* (Vol. 176, pp. 245–258). https://doi.org/10.1016/S0079-6123(09)17615-X

Spence, C., & Driver, J. (1996). Audiovisual Links in Endogenous Covert Spatial Attention. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(4), 1005–1030.

Stapells, D. R., Linden, D., Braxton Suffield, J., Hamel, G., Picton, T. W., & Wp, M. T. (1984). *Electrophysiologic Techniques in Audiology and Otology Human Auditory Steady State Potentials** (Vol. 5, Issue 2).

Thut, G., Nietzel, A., Brandt, S. A., & Pascual-Leone, A. (2006). *Behavioral/Systems/Cognitive-Band Electroencephalographic Activity over Occipital Cortex Indexes Visuospatial Attention Bias and Predicts Visual Target Detection*. https://doi.org/10.1523/JNEUROSCI.0875-06.2006

Tobimatsu, S., Zhang, Y. M., & Kato, M. (1999). Steady-state vibration somatosensory evoked potentials: physiological characteristics and tuning function. *Clinical Neurophysiology*, *110*(11), 1953–1958. https://doi.org/10.1016/S1388-2457(99)00146-7

Toffanin, P., de Jong, R., Johnson, A., & Martens, S. (2009). Using frequency tagging to quantify attentional deployment in a visual divided attention task. *International Journal of Psychophysiology*, *72*(3), 289–298. https://doi.org/10.1016/J.IJPSYCHO.2009.01.006

Tomassini, A., Ambrogioni, L., Medendorp, W. P., & Maris, E. (2017). Theta oscillations locked to intended actions rhythmically modulate perception. *ELife*, *6*. https://doi.org/10.7554/eLife.25618.001

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., & Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, *15*(1), 273–289. https://doi.org/10.1006/nimg.2001.0978

Uemura, J.-I., Hoshino, A., Go Igarashi, |, Matsui, Y., Chishima, M., Hoshiyama, | Minoru, & Mazaheri, A. (2021). Pre-stimulus alpha oscillation and post-stimulus cortical activity differ in localization between consciously perceived and missed near-threshold somatosensory stimuli. *Eur J Neurosci*. https://doi.org/10.1111/ejn.15388

Van Diepen, R. M., Cohen, M. X., Denys, D., & Mazaheri, A. (2015). Attention and temporal expectations modulate power, not phase, of ongoing alpha oscillations. *Journal of Cognitive Neuroscience*, *27*(8), 1573–1586. https://doi.org/10.1162/jocn_a_00803

Van Diepen, R. M., Foxe, J. J., & Mazaheri, A. (2019). The functional role of alpha-band activity in attentional processing: the current zeitgeist and future outlook. *Current Opinion in Psychology*, *29*, 229–238. https://doi.org/10.1016/J.COPSYC.2019.03.015

van Diepen, R. M., & Mazaheri, A. (2017). Cross-sensory modulation of alpha oscillatory activity: suppression, idling, and default resource allocation. *European Journal of Neuroscience*, *45*(11), 1431–1438. https://doi.org/10.1111/ejn.13570

Van Dijk, H., Schoffelen, J. M., Oostenveld, R., & Jensen, O. (2008). Prestimulus oscillatory activity in the alpha band predicts visual discrimination ability. *Journal of Neuroscience*, *28*(8), 1816–1823. https://doi.org/10.1523/JNEUROSCI.1853-07.2008

Worden, M. S., Foxe, J. J., Wang, N., & Simpson, G. V. (2000). *Anticipatory Biasing of Visuospatial Attention Indexed by Retinotopically Specific-Band Electroencephalography Increases over Occipital Cortex*. http://www.jneurosci.org/cgi/content/full/4016

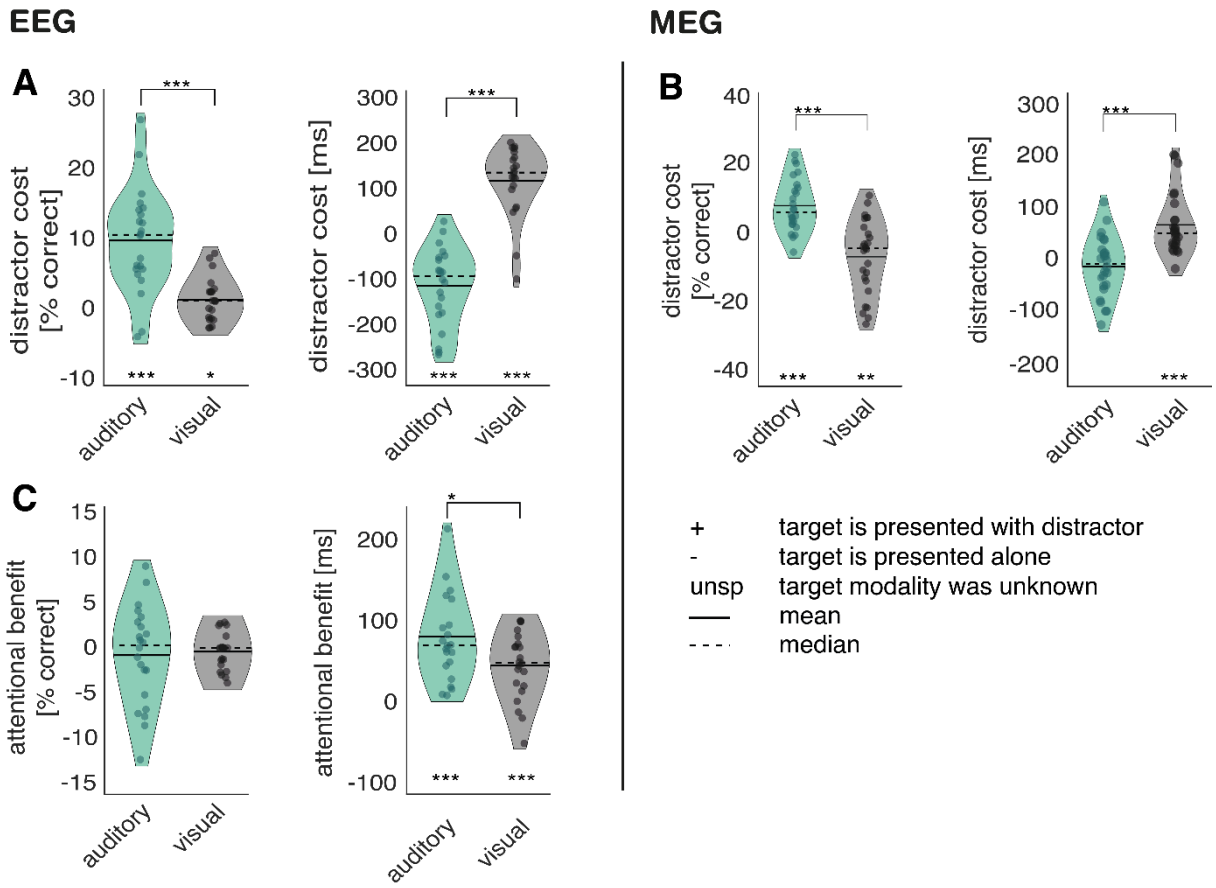
Yang, X., Fiebelkorn, I. C., Jensen, O., Knight, R. T., & Kastner, S. (2023). Differential neural mechanisms underlie cortical gating of visual spatial attention mediated by alpha-band oscillations. *BioRxiv*, 2023.08.21.553303. https://doi.org/10.1101/2023.08.21.553303

Zhigalov, A., Herring, J. D., Herpers, J., Bergmann, T. O., & Jensen, O. (2019). Probing cortical excitability using rapid frequency tagging. *NeuroImage*, *195*, 59–66. https://doi.org/10.1016/J.NEUROIMAGE.2019.03.056

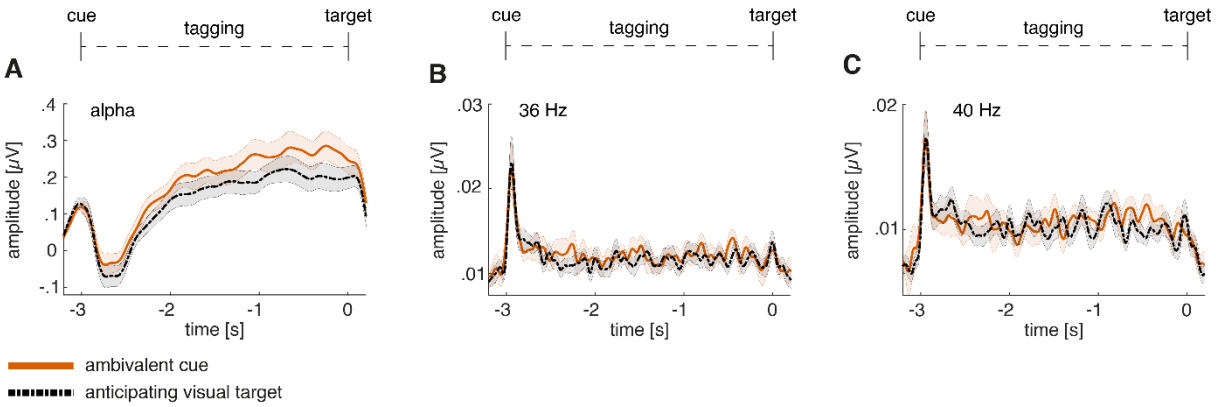
Zhigalov, A., & Jensen, O. (2020). Alpha oscillations do not implement gain control in early visual cortex but rather gating in parieto-occipital regions. *Human Brain Mapping*, *41*(18), 5176–5186. https://doi.org/10.1002/hbm.25183

Zumer, J. M., Scheeringa, R., Schoffelen, J. M., Norris, D. G., & Jensen, O. (2014). Occipital Alpha Activity during Stimulus Processing Gates the Information Flow to Object-Selective Cortex. *PLoS Biology*, *12*(10). https://doi.org/10.1371/journal.pbio.1001965

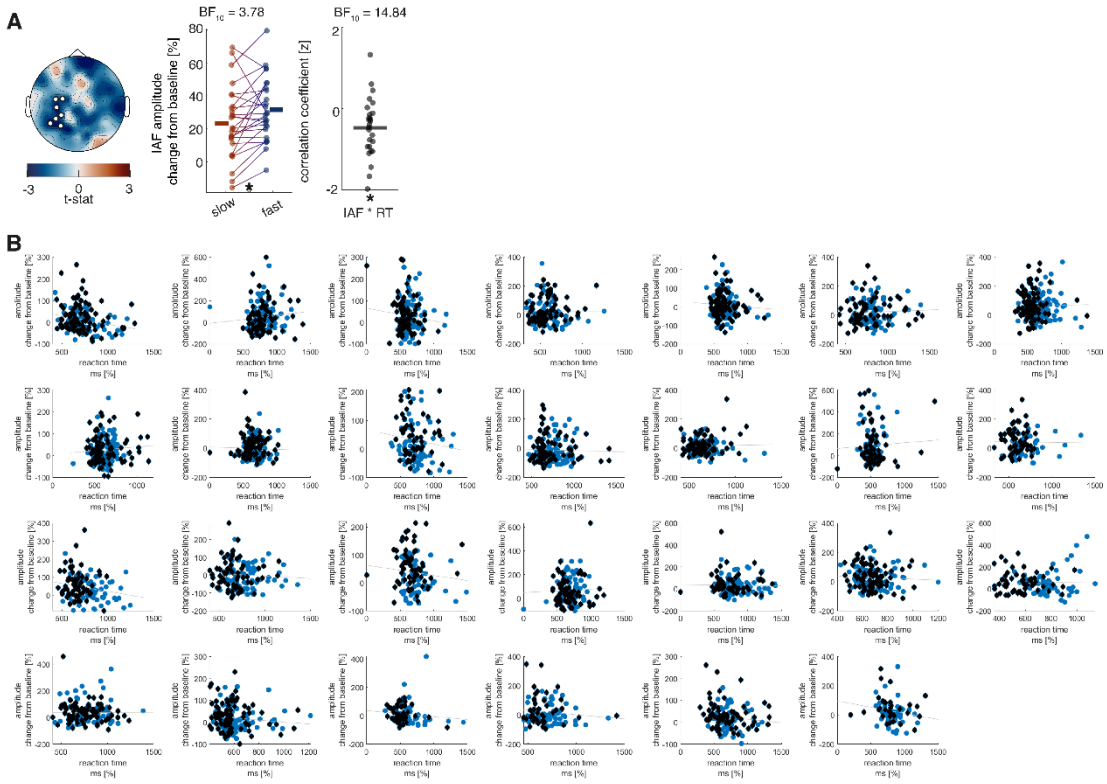
Supplementary Materials



SUPPL Fig. 1 | Distractor cost and attentional benefit. A-B, Illustration of distractor cost: mean performance over trials with distractors was subtracted from mean performance over trials without distractors. Distractor effects were observable for accuracy as well as reaction time; **A**, accuracy: auditory- – auditory +: *M* = 10.0 %; *SD* = 7.3; *p* < .001; $t_{(21)}$ = 7.32; visual- – visual+: *M* = 1.5%; *SD* = 3.06; *p* = .02). The effect was stronger for auditory than for visual target trials (*p* < .001; $t_{(21)}$ = 7.67). Reaction time: (auditory- - auditory+: *M* = -108.1 ms; *SD* = 84.8; *p* < .001; $t_{(21)}$ = -5.98; visual- - visual+: *M* = 123.6 ms; *SD* = 76.3; *p* < .001; $t_{(21)}$ = 7.60). auditory distracters decreased response time to visual targets (*p* < .001; $t_{(21)}$ = -11.99). **B**, (accuracy: auditory- – auditory +: M = 7.2 %; SD = 7.5; p = .001; t(25) = 4.9; visual- – visual+: M = -7.6%; SD = 10.80; p < .01; t(25) = -3.59; Reaction time: auditory- – auditory +: M = -20.64 ms; SD = 57.6; n.s.: p = .08; t(25) = -1.83; visual- – visual+: M = 60.1 ms ; SD = 58.52; p < .001; t(25) = 5.23). **C**, Illustration of attentional benefit: mean performance over unspecified trials was subtracted from mean performance over modality-cued trials without distractor. attentional benefit auditory: unspecifically cued auditory targets - informatively cued auditory targets = M = 81.2 ms; SD = 54.9; p < .001; t(21) = 6.94; attentional benefit visual: unspecifically cued visual targets - informatively cued visual targets -; M = 54.4 ms; SD = 41.1 ; p < .001; t(21) = 5.19). The magnitude of the effect on reaction time also differed between conditions (p = .043; t(21) = 2.16), with stronger attentional benefit for auditory target cues. Attentional cues did not affect response accuracy, neither in auditory nor visual target conditions (auditory: p = 0.49; visual: p = 0.32). EEG Study: N = 22; MEG-Study: N = 27; *** sig < .001; ** sig. < .01; * sig. < .05;

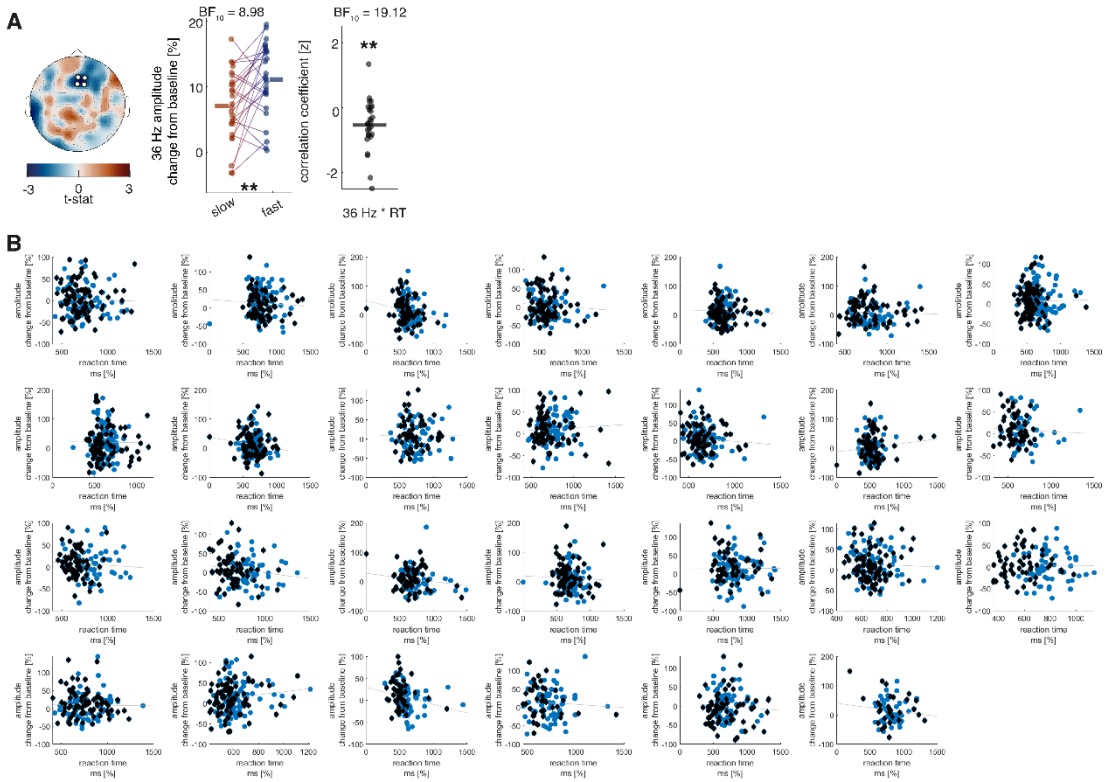


SUPPL Fig. 2 | Timecourse of alpha activity and frequency-tagging responses for the ambivalent compared to the visually-cued condition. A, alpha activity compared between expecting a visual target and having received an ambivalent cue. **B**, 36 Hz frequency-tagging response between expecting a visual target and having received an ambivalent cue. **C**, 40 Hz frequency-tagging response between expecting a visual target and having received an ambivalent cue.

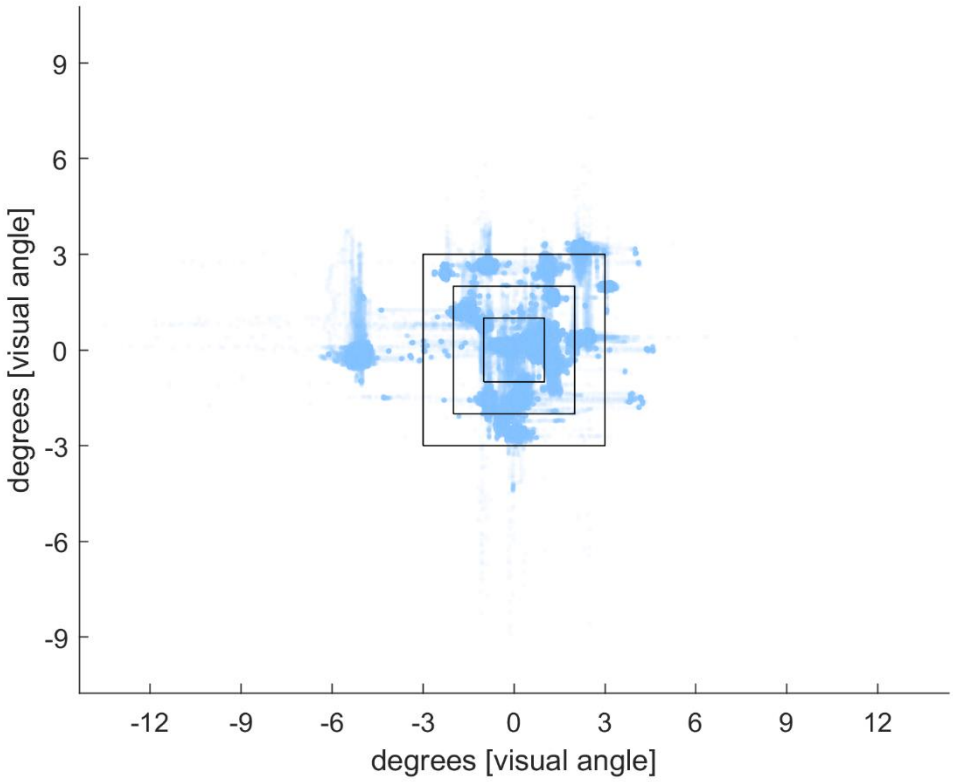


SUPPL Fig. 3 | Correlation of prestimulus alpha change from baseline with reaction time in the MEG study. A, The analysis was performed using a cluster-permutation approach, testing a correlation model against a 0-correlation model. Clusters significantly diverging from the 0-correlation model are presented topographically (p = .037). Additionally, median splits between fast and slow reaction time trials (p = .013; $t_{(25)}$ = -2.67) as well as correlation coefficients (p = .003; $t_{(25)}$ = -3.34) of these clusters are displayed for all participants. A negative correlation is visible

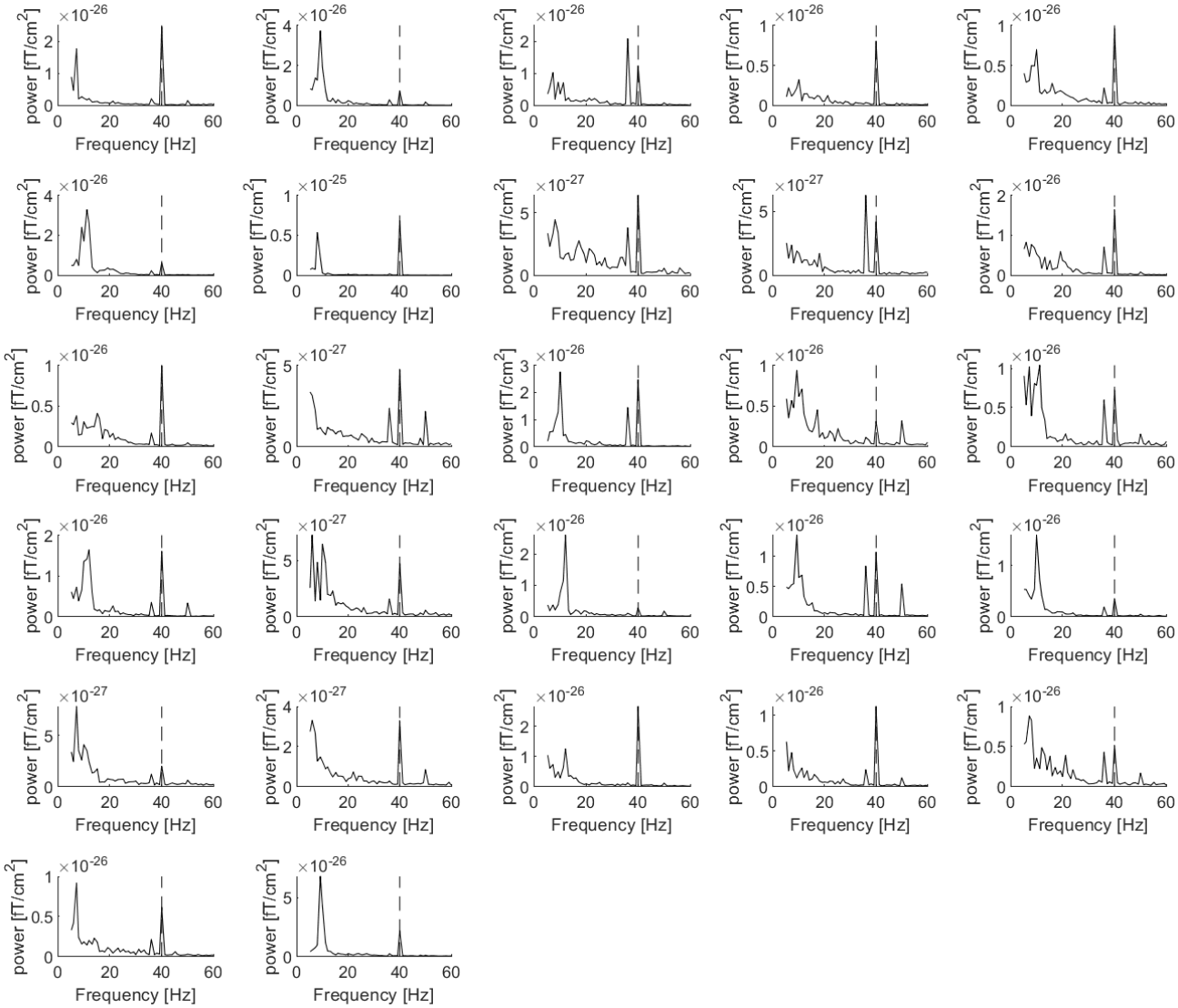
between alpha modulation and reaction times in the last 500 ms before target onset when expecting a visual target. B, Correlation between alpha modulation and reaction time for each participant. Black diamonds represent trials from the first block (without distractor) and blue dots represent trials from the second block (with auditory distractor).



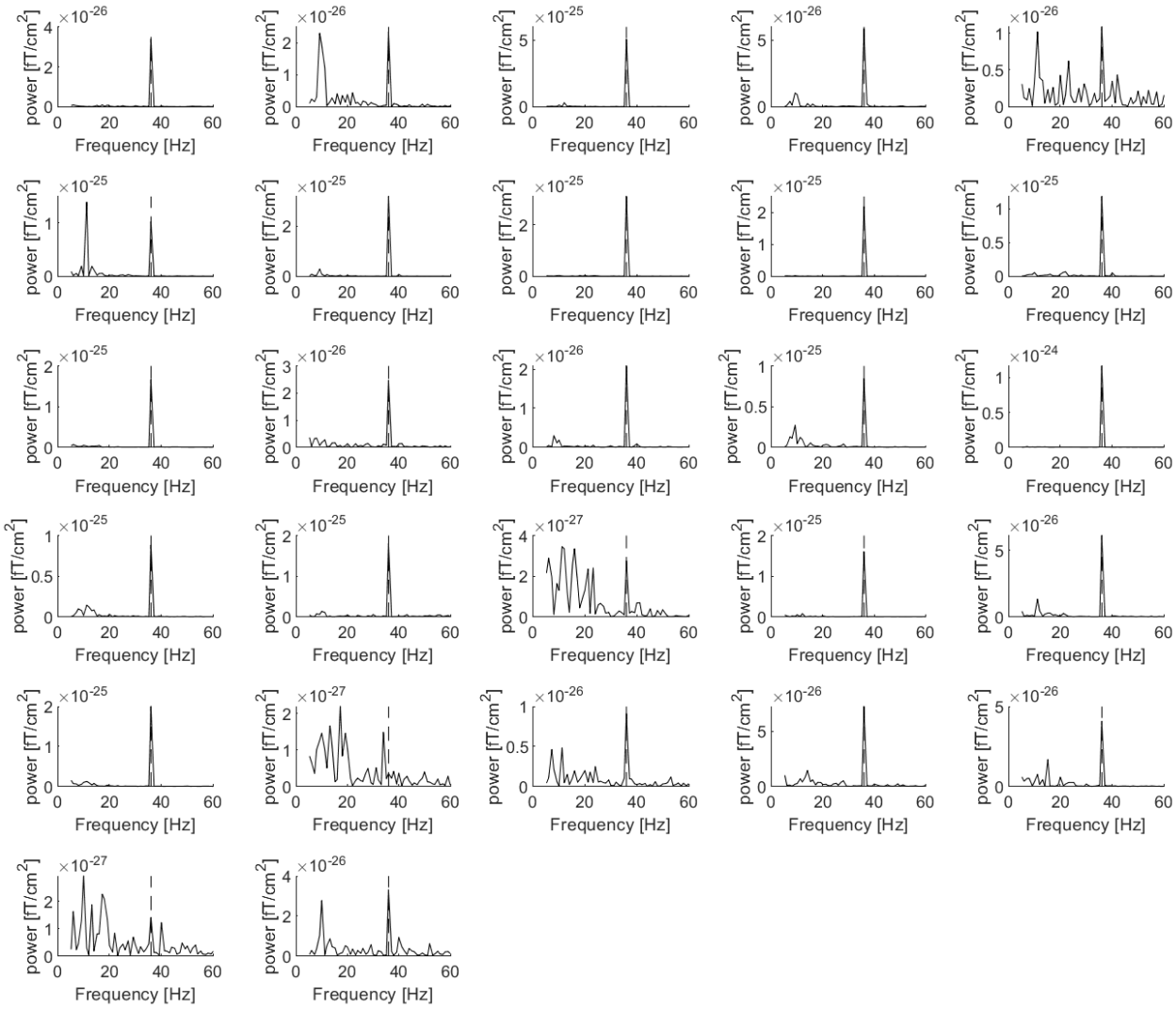
SUPPL Fig. 4 | Correlation of 36 Hz change from baseline with reaction time in the MEG study. A, The analysis was performed using a cluster-permutation approach, testing a correlation model against a 0-correlation model. Clusters significantly diverging from the 0-correlation model are presented topographically (p = .040). Additionally, median splits between fast and slow reaction time trials (p = .005; $t_{(25)}$ = -3.10) as well as correlation coefficients (p = .002; $t_{(25)}$ = -3.46) of these clusters are displayed for all participants. A negative correlation is visible between 36 Hz modulation and reaction times in the last 500 ms before target onset when expecting a visual target. B, Correlation between 36 Hz modulation and reaction time for each participant. Black diamonds represent trials from the first block (without distractor) and blue dots represent trials from the second block (with auditory distractor).



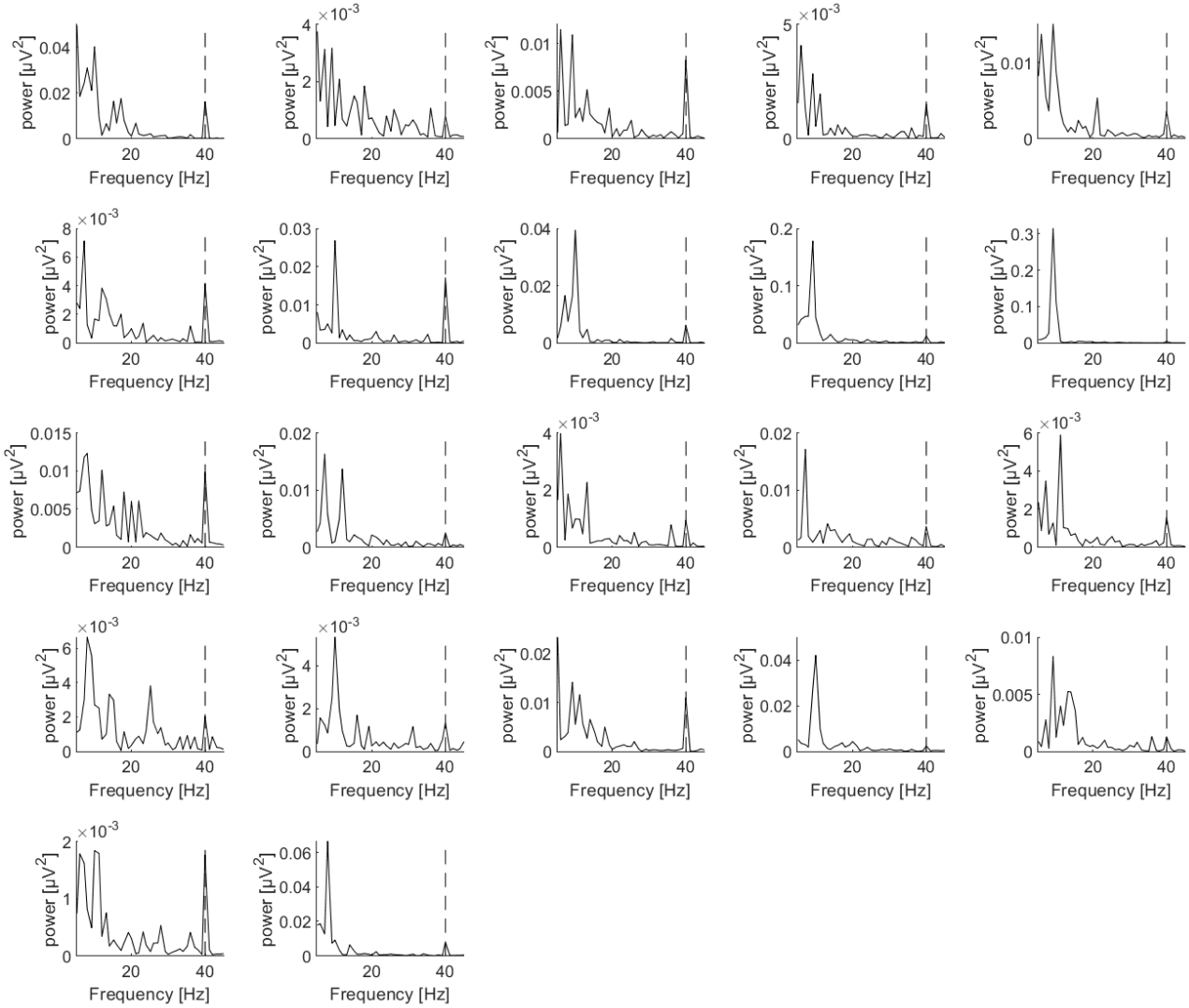
SUPPL Fig. 4 | Illustration of eye-tracking during the cue-to-target interval (2.5 – 0 s before target onset). All datapoints of eye-positions during the cue-to-trial interval for all trials and all participants were plotted with 5% visibility on top of each other. Only 3% of datapoints showed eye-movement larger than 3 degrees of visual angle away from the fixation cross.



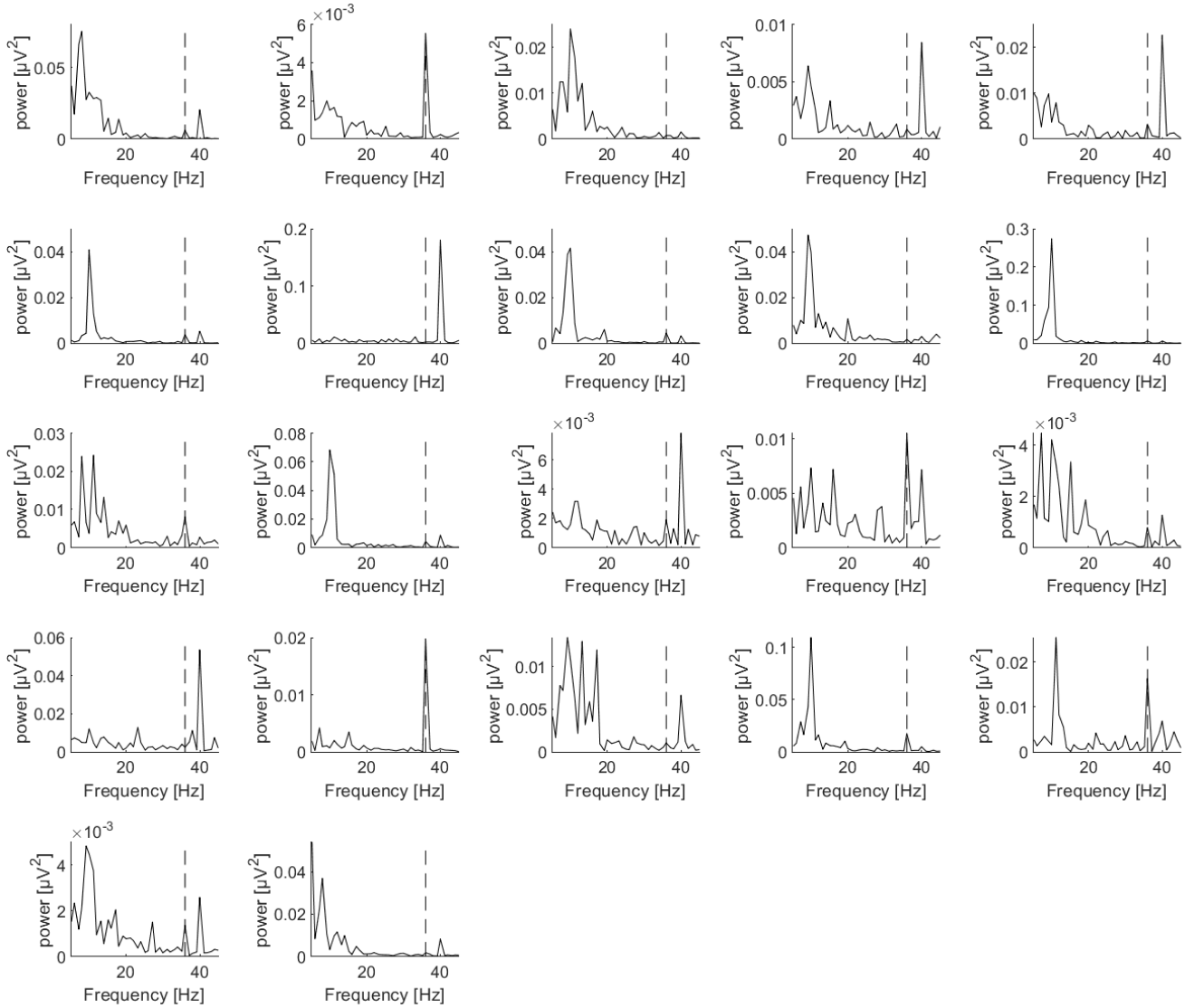
SUPPL Fig. 6 | Individual ERP power spectra of the cue-to-target interval when anticipating an auditory target in the MEG-study. Fast-fourier transformation was applied to the averaged trials using a dynamic hanning-tapered sliding time-window of 7 cycles per frequency. The Dotted line represents 40 Hz (auditory frequency-tagging).



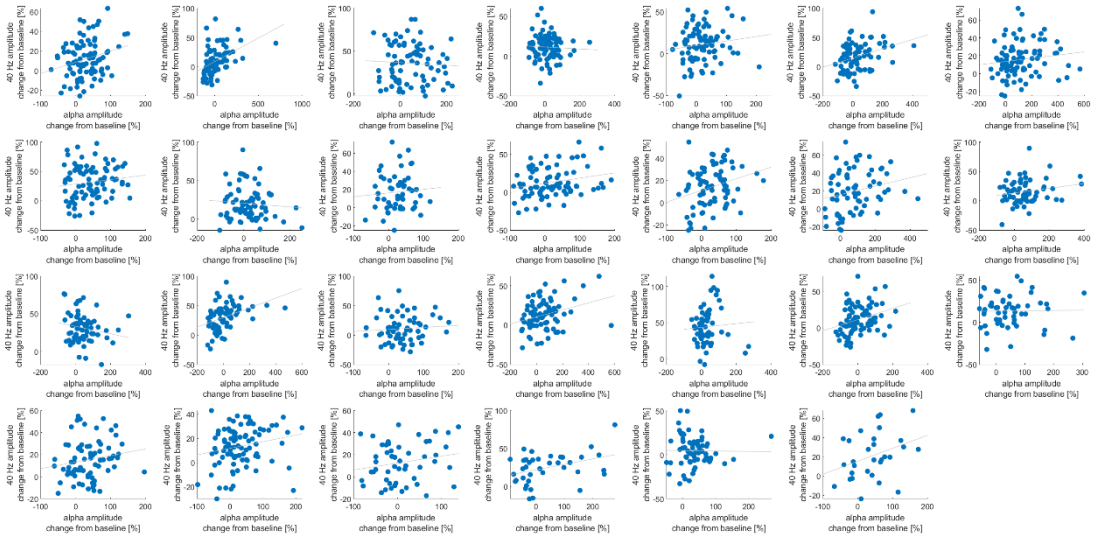
SUPPL Fig. 7 | Individual ERP power spectra of the cue-to-target interval when anticipating a visual target in the MEG-study. Fast-fourier transformation was applied to the averaged trials using a dynamic hanning-tapered sliding time-window of 7 cycles per frequency. The Dotted line represents 36 Hz (auditory frequency-tagging).



SUPPL Fig. 8 | Individual ERP power spectra of the cue-to-target interval when anticipating an auditory target in the EEG-study. Fast-fourier transformation was applied to the averaged trials using a dynamic hanning-tapered sliding time-window of 7 cycles per frequency. The Dotted line represents 40 Hz (auditory frequency-tagging).



SUPPL Fig. 9 | Individual ERP power spectra of the cue-to-target interval when anticipating a visual target in the EEG-study. Fast-fourier transformation was applied to the averaged trials using a dynamic hanning-tapered sliding time-window of 7 cycles per frequency. The Dotted line represents 36 Hz (auditory frequency-tagging).



SUPPL Fig. 10 | Exemplary illustration of the correlation between alpha power (0.5 – 0 s before target onset) and 36 Hz steady-state response (0.5 – 0 s before target onset) for each participant. Alpha activity was averaged over the significant group difference cluster for alpha condition differences (seed cluster). Frequency-tagging activity was averaged over the significant cluster in the correlation with the alpha seed activity.