Temporal binding of Visual and Motor cortical oscillations revealed by BOLD response in Humans

Alessandro Benedetto[1], Paola Binda[1], Mauro Costagli[2,3], Michela Tosetti[2,3] & Maria Concetta Morrone[1*]

[1]Department of Translational Research on New Technologies in Medicine and Surgery, University of Pisa, Pisa, Italy; [2]IRCCS Stella Maris, Calambrone, Pisa, Italy; [3]Imago 7 Research Foundation, Pisa, Italy

*Correspondence: concetta.morrone@unipi.it

Summary

Action and perception need to be coordinated continuously over time, and neural oscillations may be instrumental in achieving such synchronization. Here we demonstrate that behavioral visual discrimination and the BOLD activity of V1 oscillates rhythmically in the theta range (around 5 Hz), synchronized to motor action (button press). The oscillations are present in V1 even when participants do not make a visual discrimination, suggesting an automatic modulation in synchrony with action onset. The amplitude of the oscillation in V1 is predicted by the activity in M1 before action onset, and functional connectivity between V1 and M1 change as a function of stimulus-delay. The results are well modelled by considering that V1 BOLD is modulated by preparatory motor signal and by rhythmic gain modulation in phase with action onset. They suggest that synchronous oscillatory activity between V1 and M1 mediates the strong temporal binding fundamental for active visual perception.

Keywords: visuo-motor integration, V1, neural endogenous oscillations, behavioral oscillations, BOLD, action, functional connectivity, visual discrimination, 7T fMRI.

Acknowledgment

This project was supported by the European Research Council (ERC) under the European Union's Horizon 2020 program (Grant Agreement No 832813-GenPercept and No 801715-PUPILTRAITS) and by MIUR - PRIN 2017 - grant 2017SBCPZY_02. We thank Eleanor Reynolds and David Burr for proofreading the manuscript. The authors would like to thank Dr. Brian Burns (GE Healthcare) for his implementation of the 3D-MP2RAGE sequence on the MR950 system.

INTRODUCTION

For an efficient interaction with the external world, our brain has to continuously integrate sensorimotor information, with a tight temporal coupling between motor and visual activity. This leads to the idea that sensory processing is not a passive, static process but, rather, an active, dynamic process where not only the motor system informs the visual brain of the upcoming events (Helmholtz, 1866; Sperry, 1950; von Holst and Mittelstaedt, 1950; for a review see Wurtz, 2018), but also vision may dynamically change motor activity (Gallese et al., 1996; Goodale et al., 1986; for a review see Goodale and Milner, 2013). The process is inherently very complex because it has to adjust dynamically to take into account the sensor and effector movements in both external and body coordinates.

The continuous movement of our receptors in the environment gives the great advantage of efficiently detecting changes in the world but it imposes heavy costs: perceptual information acquired during movement is often noisy and potentially misleading, requiring continuous temporal binding between the flow of motor and sensory analysis. Eye movements represent a paradigmatic case of active sensing (Rao and Ballard, 1999). Visual perception around the time of saccadic execution is suppressed to cancel out spurious visual signals generated by the fast rotation of the eyeball (Burr et al., 1994), and visual receptive fields are predictively remapped to mediate perceptual stability across eye movements (Duhamel et al., 1992; for a review see: Binda and Morrone, 2018; Ross et al., 2001; Wurtz, 2018). However, the problem is not limited to eye movements, but generalizes to many other voluntary actions, such as reaching or pointing movements, although the visual consequences are less obvious than with eye movements (Blakemore et al., 1998; Wolpert et al., 1995). For example, visual time perception is altered during saccades (Morrone et al., 2005), but also during hand movement (Merchant and Yarrow, 2016; Tomassini et al., 2018; Yokosaka et al., 2015). Similarly, the motor-induced suppression of perception, very strong around saccadic onset, occurs also for other actions and senses, indicating the existence of a pervasive sensorimotor network in the brain that optimizes sensory processing around the time of action (Melloni et al., 2009; Schroeder et al., 2010).

The problem of visual-motor integration becomes particularly complex when analyzing the constraints that the brain must face to solve the temporal coordination across motor and sensory processing. The flow of temporal information for each element of the sensory-motor network has different processing times, speeds and delays, which can vary widely even within modalities, depending on the features of the sensory stimulation or the pattern of the programmed movement. The director of the time coordination in the brain has been thought to reside in the motor system. The master-clock that should provide a unitary sense of time and coordinate the various processes is consistently found to be associated with motor cortex or motor networks (Coull et al., 2004; Merchant and Yarrow, 2016; Nobre and O'Reilly, 2004). Furthermore, programming an action dramatically modulates visual perception over long time. Also fine and basic discrimination sensitivities thought to be mediated by primary visual cortex such as contrast increment detection or orientation discrimination (Benedetto et al., 2016; Tomassini et al., 2015; Zhang et al., 2019) can oscillate over a very long time, in the order of seconds and also before action onset (during motor preparation phase). Importantly,

the modulations are rhythmic, in the theta or alpha frequency ranges (Benedetto et al., 2018; Benedetto et al., 2016; Tomassini et al., 2017; Tomassini et al., 2015; Zhang et al., 2019). The fact that behavior is rhythmically modulated in synchrony with action strongly suggests that endogenous neuronal rhythms may play an important role in sensory-motor temporal coupling.

Coordination and communication among distal brain areas might be easier if orchestrated by synchronous neuronal oscillation, rather than by an abrupt signal (Canolty et al., 2006; Fries, 2015; Roelfsema et al., 1997). Fiebelkorn and Kastner (2019) have recently demonstrated in monkeys that the phase of theta rhythms, that are widely diffuse in the brain, may absolve the synchronization task by altering the functional connectivity between brain structures. Also in humans, functional connectivity between occipital and parietal brain areas is modulated in synchrony with theta rhythm phase (Hanslmayr et al., 2013) . Interestingly, the phase and power of theta neuronal oscillations predict visual discrimination performance (Kaplan et al., 2012; Tomassini et al., 2017) around the time of execution of a voluntary action in humans.

Capitalizing on our previous evidence that a simple action such as a button press can reveal a theta rhythmic modulation of basic visual discrimination (Benedetto et al., 2016; Zhang et al., 2019), we study the neuronal circuitry mediating the effect. Using a 7T scanner to measure BOLD visual and motor cortical responses during a visual discrimination task, we show that the rhythmic modulation is present as early as the primary visual cortex, and that the visual cortical rhythms are shared with the motor and somatosensory cortex. Crucially, the magnitude of the effect of the modulation across individuals is associated with the activity in the motor cortex during motor preparation. The data suggest that motor and visual cortex BOLD activity are synchronized by a common rhythm, and that the state of pre-stimulus activity is predictive of the BOLD and behavioral response to the visual event.

RESULTS

Behavioral oscillation

Using psychophysical techniques, we demonstrate that the accuracy of discriminating the spatial frequency of a briefly presented grating oscillates rhythmically in synchrony with a voluntary action onset, as previously shown for orientation and contrast discrimination (Benedetto et al., 2016; Tomassini et al., 2015; Zhang et al., 2019). Figure 1A (black symbols) shows how the accuracy in the spatial frequency discrimination between the superior and inferior part of a grating (see insert in Fig 1) varies as a function of the delay between the action (keypress) to start the trial, and the presentation of a brief grating stimulus (data averaged over 7 participants). The accuracy timeseries is best fitted by a sinusoidal waveform with a highly significant peak at 5.4 Hz (R^2 = 0.427, p = 0.005, corrected for multiple comparisons from surrogate data analysis: see methods). Interestingly, the first minimum of the sinusoidal fit occurs at about 70 ms from action onset, which corresponds to the well-known phenomenon of motor-induced suppression (Aliu et al., 2009).

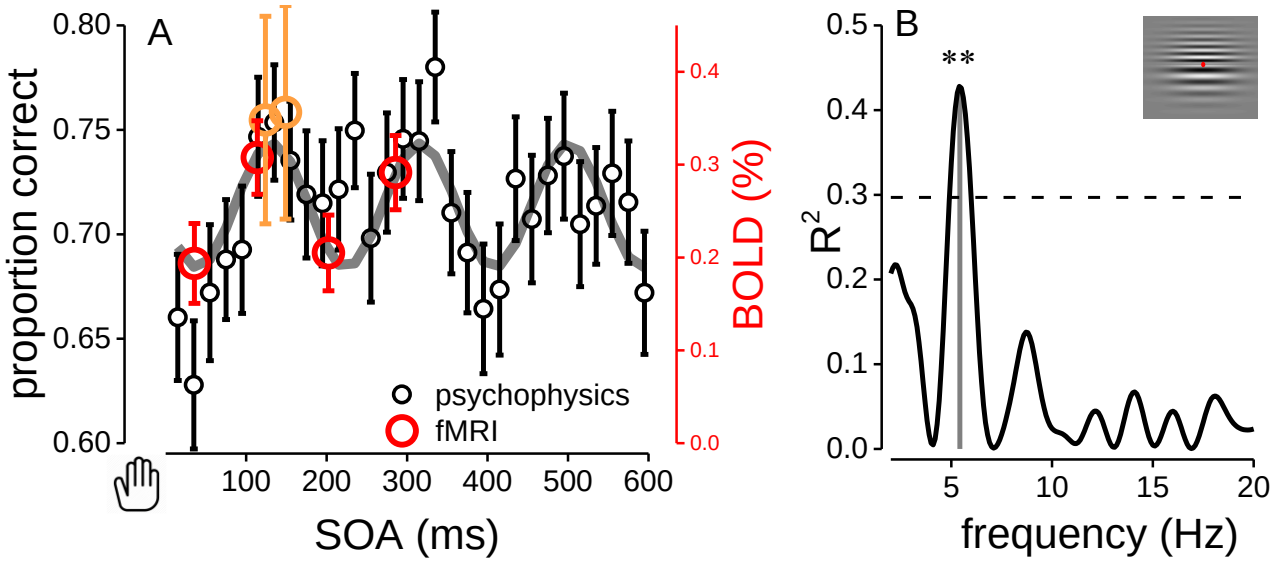

Figure 1. Behavioral and BOLD oscillations. A: spatial frequency discrimination accuracy (black symbols: mean ± s.e.m.; aggregate observer, N = 7) as a function of stimulus onset asynchrony (SOA) with the best sinusoidal fit of the accuracy time-course (gray line; 5.4 Hz). % BOLD signal change in V1 from figure 3 at the six SOAs tested in the scan (red and orange circles: mean ± s.e.m.; N = 21 except N = 17 for 310 ms, and N = 5 for the orange circles). **B**: goodness of fit of a sinusoidal waveforms with frequency parameter fixed between 3 and 20 Hz for the accuracy in A (black lines) show a strong and significant peak at 5.4 Hz (R^2 = 0.427) well above the 95th percentile of the R^2 distribution obtained from permuted dataset (dashed line). Asterisks: statistical significance 0.01 > ** > 0.001.

BOLD visual cortex oscillation

We measured the BOLD response to the same stimuli and action used for the psychophysics study in an ultra-high magnetic field (7T) scanner. Although the stimuli were very brief (33 ms), the BOLD response along the calcarine sulcus was strong and reliable, both when the stimuli were presented after the keypress action (Figure

2C) and when presented in a more typical condition without the button press (Figure 2B). However, the timecourses of the two responses were different (Figure 2D). Without the action at the start of the trial, the BOLD response followed the typical hemodynamic response of V1 (Figure 2D, black triangles); for stimuli presented after the keypress, the BOLD response reached a similar peak response, but then attenuated rapidly (Figure 2D, red squares). The major differences between the two conditions are between 6 and 12 s. Therefore, we quantified the responses to vision and action in this interval.

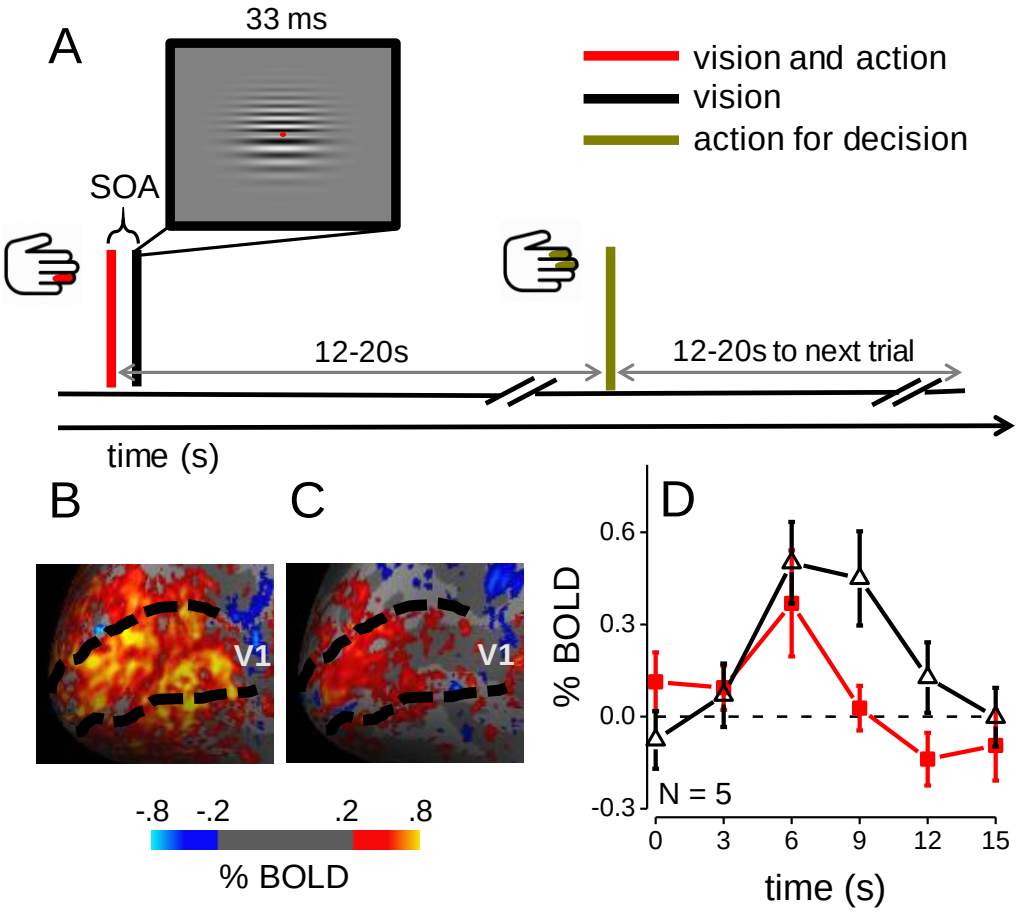

Figure 2. Procedure and BOLD response in V1. A: schematics of the event-related design measuring BOLD activity in response to visual stimuli preceded by a voluntary action (*vision and action*) and to voluntary actions used to report the perceptual decision (*action for decision*). In separate sessions, we also measured BOLD activity in response to visual stimuli alone (not shown). **B-C**: BOLD response (GLM beta weights at delay 6 s, representing the peak of the hemodynamic response) to visual stimuli presented alone (B) or to *vision and action* events (C), computed after projecting all participants' data to a common mirror-symmetric template of the cortical surface and averaging data across hemispheres. Dashed lines report the borders between V1 and V2. **D**: GLM beta weights estimated for the V1 subregion representing the stimulus area to visual stimuli presented alone (black triangles) and to *vision and action* events (red squares). N = 5, bars s.e.m.

To investigated whether the BOLD response oscillates over time, we varied the delay between the visual stimulus presentation and the action onset, like in the psychophysics experiment reported in Figure 1A. Figure 3A shows the BOLD responses of V1 for four SOAs from action-onset. The responses for SOAs of 70 and

150 ms are clearly different for much of the timecouse (Figure 3A, blue and dark green curves). Although the curves are similar (not significantly different) at 0 and 3 secs, they are clearly displaced between 6 and 12 s (see modelling section further discussion about the dynamics). To quantify the effect, we averaged BOLD responses from 6 to 12 s for the 4 SOAs (Figure 3B). The effect is cyclical with SOA, with responses lower at 70 and 230 ms than 150 and 310 ms. Linear mixed-effect model analysis shows a significant effect of SOA (see Table 1), with post-hoc paired t-tests demonstrating a significant difference of 150 ms and 310 ms versus 70 and 230 ms SOA responses (see table 1, Model Comparison column), consistent with an oscillatory modulation at theta range similar to that observed in Figure 1 for discrimination accuracy. The scatter plots show the effect is present in most of the individual subjects, despite the large variability in the BOLD response across the sample. Interestingly, the BOLD responses at 150 and 310 ms SOAs (Figure 3F) and at 70 and 230 ms SOAs (not shown) are practically equal in all subjects. However, the average BOLD response at 150 ms and 310 ms is higher for almost all subjects compared to the average BOLD response at 70 and 230 ms SOA (Fig 3C). The results of figure 3D-F are not consistent with a form of transient motor-suppression, which would predict a monotonically increasing response at larger SOAs (Aliu et al., 2009). Instead, this data suggests that BOLD activity in V1 is rhythmically modulated by the delay between the start of the voluntary action and the appearance of the visual stimuli.

To compare the BOLD data with the behavioral data collected outside the scanner, we plotted the two sets of data (with different abscissae) in superposition in Figure 1A. The red symbols plot the data of Figure 2B, the orange symbols data obtained from another cohort of subjects (N = 5) where responses to 150 and 180 ms SOAs were measured. The agreement between all sets of results is very strong with the grey curve fitting the BOLD data well (R^2 = 0.485; amplitude free parameter): the temporal minima and the maxima of BOLD responses coincide with the minima and maxima of the spatial frequency discrimination performance, suggesting the presence of similar rhythmical BOLD and behavioral sensitivity modulation at about 5 Hz.

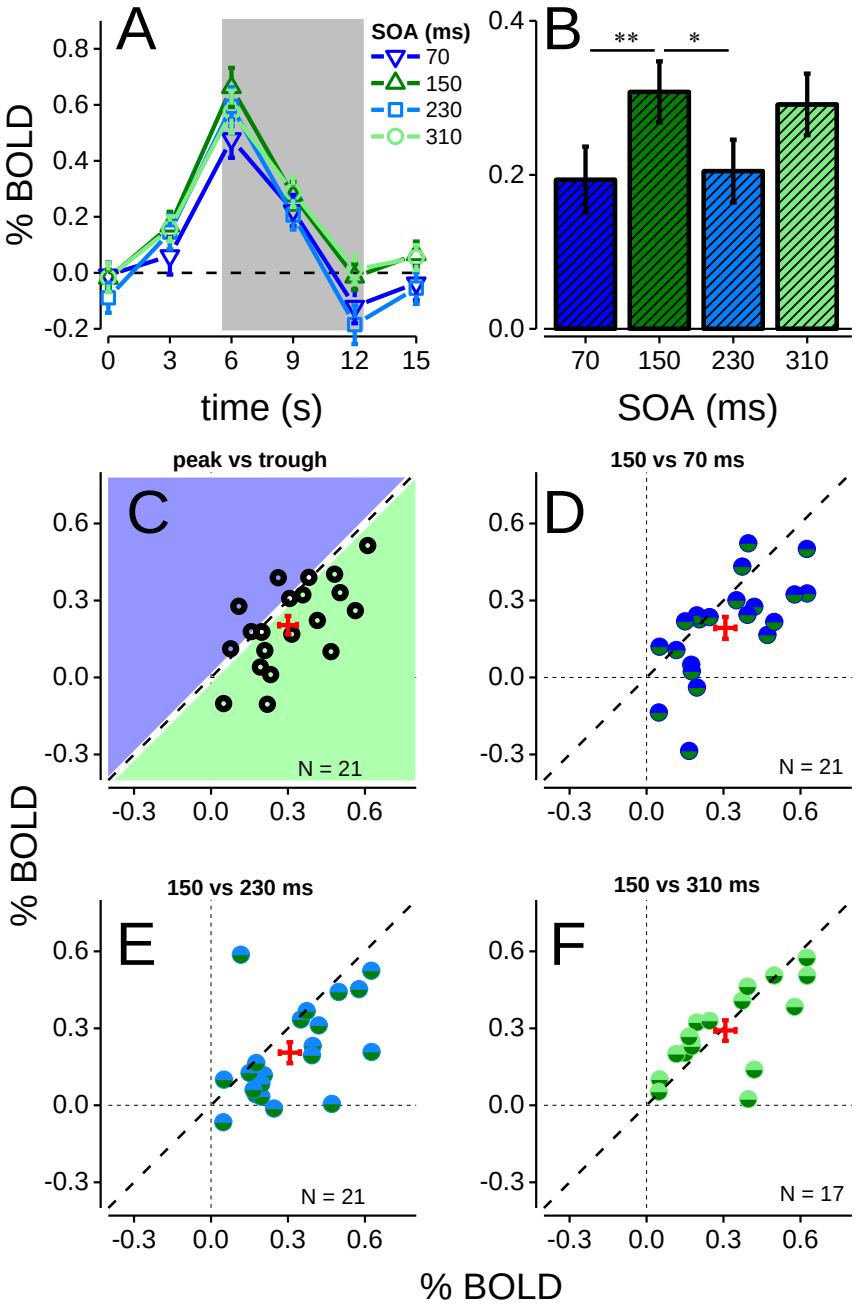

Figure 3. Temporal oscillation in V1 BOLD responses and comparison across different stimulus-action delays. A: Timecourses of the GLM beta weights (mean and s.e.m.) representing the BOLD response to *vision and action* events in the V1 subregion for different SOAs from button press (down triangle: 70 ms, up-triangle: 150 ms; square: 230 ms; circle: 310 ms SOA) averaged across subjects. **B**: GLM beta weights averaged from 6 to 12 s, indicated by the gray area in A, for the different SOAs. Asterisks: statistical significance across subjects (0.05 > * > 0.01 > ** > 0.001) of Paired Samples Student's t-test (70 vs 150: t(20) = -3.374, p = 0.003; 150 vs 230: t(20) = 2.595, p = 0.017). **C:** Scatter plots of GLM beta weights as in B when combining events at 150 and 310 ms SOA (peak of the oscillation) for the abscissa and that the events at 70 and 230 ms SOA (trough of the oscillation) for the ordinate. **D-E-F**: as C for the 150 ms versus the 70 ms (D), 230 ms (E), 310 ms (F) SOAs. Each dot represents a single subject; red crosses mean and s.e.m.. See Table 1 for linear mixed-effect model and post-hoc test details.

Note that the BOLD modulation is present for visuomotor SOAs that usually are very difficult to distinguish, given that their difference is less than the temporal discriminability threshold for vision and action (around 150 ms) (Stetson et al., 2006).

The SOAs modulation is also apparent in the BOLD activity map of V1 (Figure 4A). The maps for 150 and 310 ms SOAs show high values, while only a few vertices show a strong BOLD response for the other two SOA conditions, especially for the 70 ms SOA. The BOLD modulation across conditions occurs at all V1 eccentricities, perhaps stronger around 2 deg, the most useful eccentricity to perform the discrimination task (Figure 4C). Interestingly, there is also a small difference between SOAs at eccentricities beyond the stimulus eccentricity (Figure 4B shows the stimulus contrast variation over eccentricity), implying that the modulatory effect is independent of the stimulus characteristics and probably related to an automatic modulation pervasive in all V1 (see Model Section).

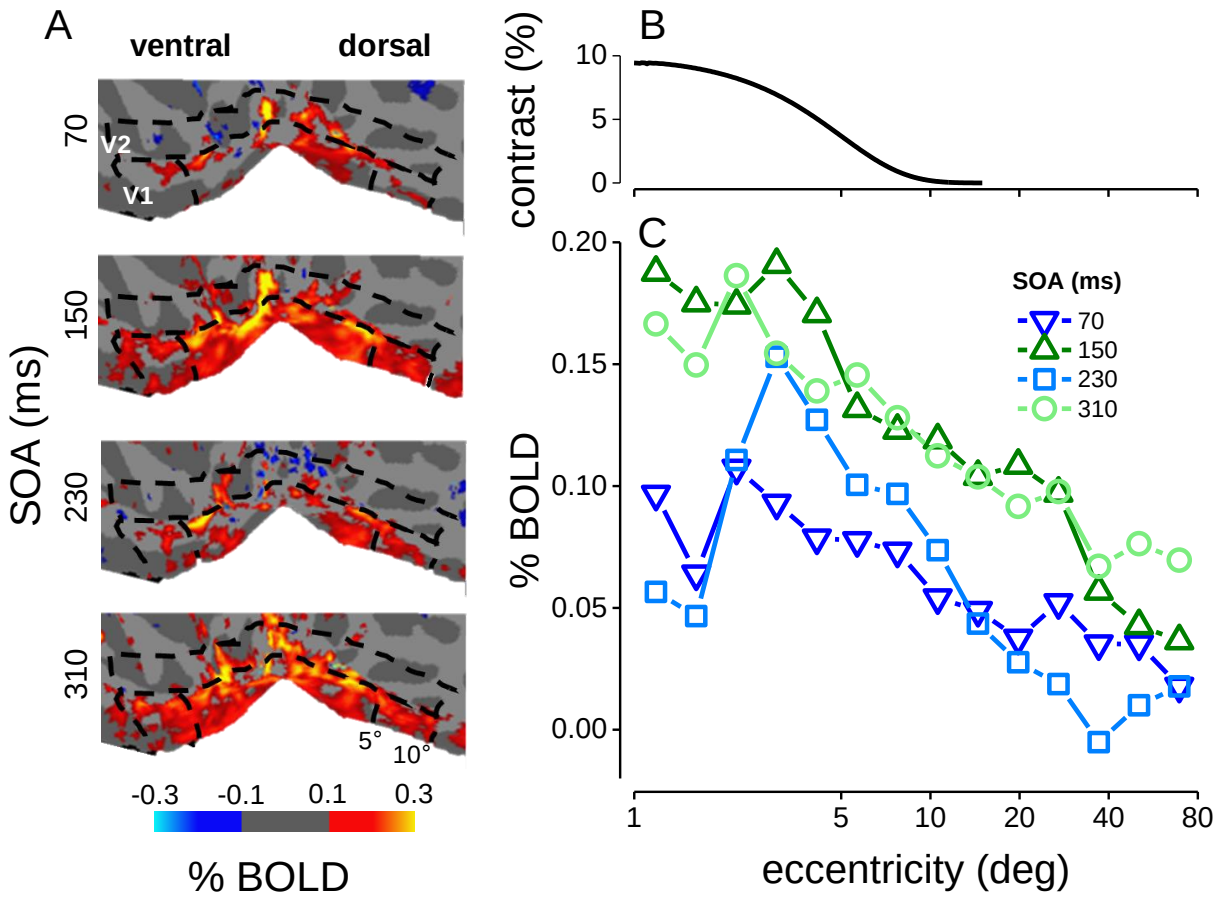

Figure 4. Effect of eccentricity on BOLD oscillations. A: BOLD response averaged in the 6-12 s time-window projected on the mirror-symmetric template of the cortical surface of FSL (see Methods). Dashed lines indicate the borders between V1 and V2 and visual eccentricities extracted from Benson et al. (2012). Activity for 150 (second panel from top) and 310 ms (bottom panel) SOAs is stronger than for 70 (first panel) or 230 (third panel from top) ms SOAs. **B**: stimulus contrast variation over eccentricity. **C**: BOLD response as in A, plotted in 14 non-overlapping logarithmic steps (1 deg to 80 deg) of eccentricity in V1 for the four SOAs (symbols and color as in Figure 3).

Motor and somatosensory cortex BOLD oscillation

Having established a rhythmic modulation of BOLD response in V1, we checked whether a similar oscillation may also occur in the motor or somatosensory cortex. For each individual subject, we defined the ROIs of motor (M1) and somatosensory (S1) cortex from the BOLD response to the action performed to report the perceptual decision of the discrimination task (from now on referred to as *action for decision* event, see Figure 2A). The M1 and S1 ROIs were located close to the Rolandic sulcus with distinct foci of activity (see example maps and ROI in supplementary Figure 1S). The overall BOLD response in M1 and S1 to the *vision and action* event (see Figure 2A) has a different hemodynamic from that observed in V1: the response is anticipated, reflecting the readiness potential during the preparation phase of the action (compare Fig 5A with Fig 5D-G; Cunnington et al., 2003). The other difference is a prolonged negative post-stimulus undershoot, much less salient in V1 (see also Figure 2).The panels on the right show how the average BOLD (between 6 and 12 s) is modulated by SOAs when the subject had to report the discrimination decision by keypress (*with action for decision*; panels B-E-F) and when not (*no action for decision*; panels C-F-I), superimposed to the best sinusoidal fit of the behavioral data of Figure 1 (see also Figure S2 and S3 in Supplementary).

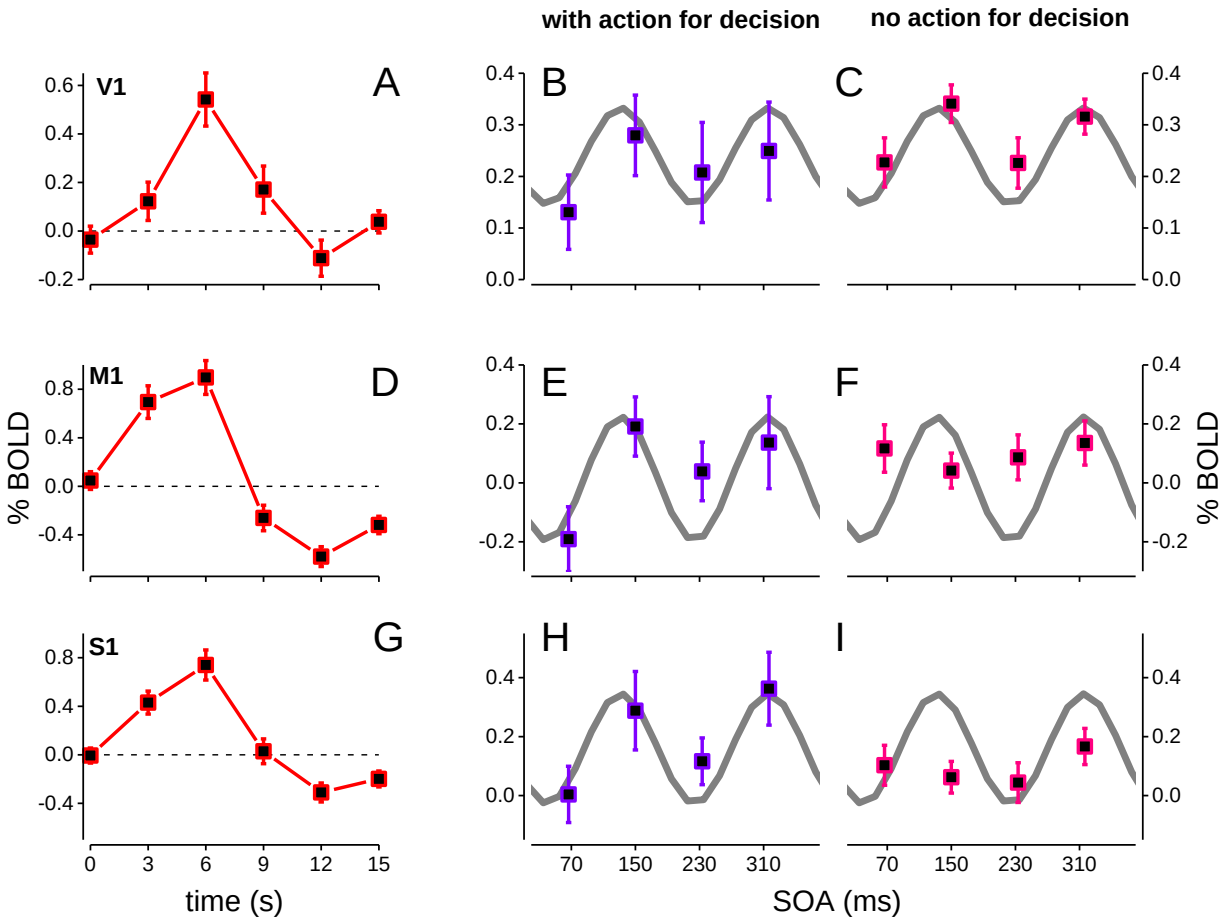

Figure 5. Effect of SOA in V1, M1 and S1 when participants did or did not make an action to report the perceptual decision. A-D-G: % BOLD response timecourse to *vision and action* events in V1 (A), M1 (D), and S1 (G), averaged across all trials and conditions. **B-E-H**: average BOLD response in the 6-12 s time-window from the keypress in V1 (B), M1 (E), and S1 (H) when the subjects had to report the perceptual

discrimination decision with an action (*with action for decision*). **C-F-I**: same as in panel B-E-H but when the subjects did not report the perceptual decision with an action (*no action for decision*), for V1 (C), M1 (F), and S1 (I). For M1 and S1 we found an SOAxTASK interaction ($p < 0.05$), resulting in a significant effect of SOA only for the task requiring the action to report the perceptual decision ($p < 0.05$). See Table 1 for linear mixed-effect model results.

V1 activity did not change significantly with participants performing the discrimination task, but both M1 and S1 did. When participants had to report the perceptual decision with a motor act, BOLD response of M1 and S1 changed rhythmically with SOA, in a similar pattern observed for V1 (see linear mixed-effect model analysis and post-hoc test in Table 1). When no motor response was required, there was no BOLD modulation with SOAs. Although the participant was not required to note the time of appearance of the stimuli, but only the difference in spatial frequency, an effect of the implicit memory of the visual stimuli timing was embedded in the BOLD modulation of motor and somatosensory cortices.

BOLD correlation analysis in the sensory-motor network

Given that a similar BOLD modulation with SOA was observed in V1, M1 and S1 (Figure 5B, E and H), we searched for a common cortical rhythm by studying the shared variability of the individual trial responses across the subject population. To this end we could not simply correlate the BOLD activity, given that our experimental design does not allow the separation between motor and visual response (SOA less than 310 ms). However, the residuals might have informative variability, sufficient to highlight common noise and functional connectivity. Figure 6A shows that the correlation of the residuals from the GLM best fit of V1 and M1 is significantly enhanced for SOAs associated with the maximum BOLD responses (150 and 310 ms) compared with the other SOAs (150 vs 70 ms: $p = 0.004$; 150 vs 230 ms: $p = 0.017$). The enhanced correlation at SOA 150 ms is not simply a by-product of higher BOLD activity in these conditions, given that the stimulus responses have been accounted for by the GLM fit. No significant modulation of correlation was observed for V1-S1 or M1-S1 ($p > 0.05$). The correlation patterns indicate that shared noise, reflecting the background signals across areas, changes rhythmically over time, being higher at the point of maximum behavioral performance. This result suggests that a common oscillatory excitability reverberates in concert in the visual and motor areas.

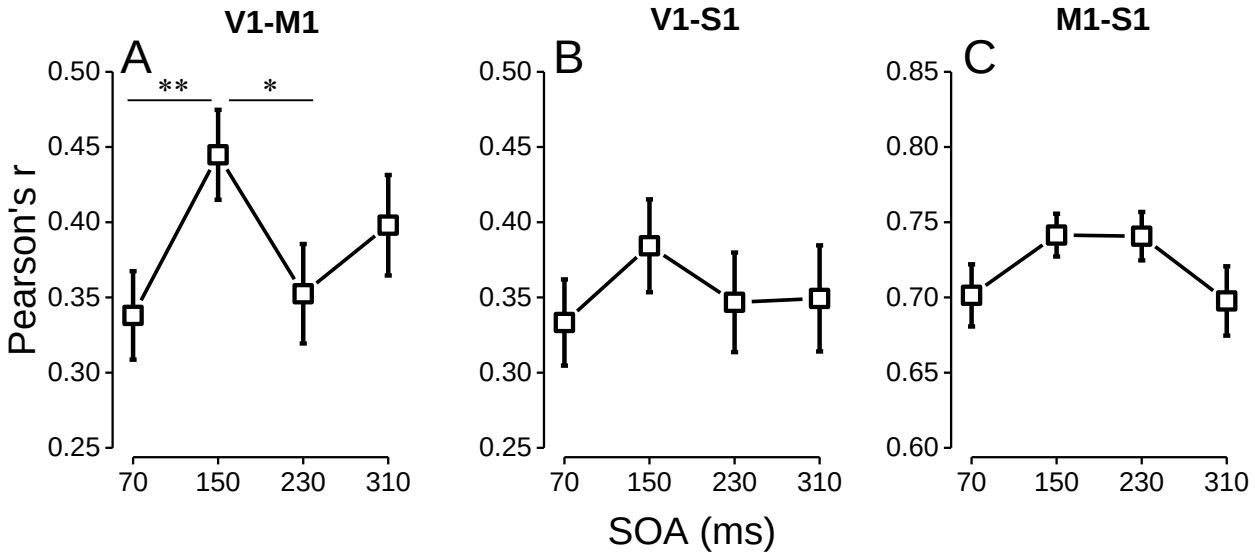

Figure 6. Cortical area correlations for the various stimulus offsets. A: coefficient of correlation (Pearson's r) between the averaged BOLD residuals across voxels of V1 and M1 as a function of SOA. Mean and s.e.m from the bootstrapped aggregate observer. The correlation between V1 and M1 is higher when stimuli are presented 150 ms and 310 ms from button press. Asterisks: statistical significance 0.05 > * > 0.01 > ** > 0.001; see Methods for statistical analysis details. **B-C**: same as in A, for V1-S1 (B) and M1-S1 (C). In these cases, the differences among correlation coefficients were not statistically significant (p > 0.05).

Is this oscillation independent of the neural activity directly evoked by the visual stimulus, and related to the pattern of the endogenous oscillations (as postulated by many researchers; for a review see VanRullen, 2016)? To answer these questions, we correlated the strength of the SOA modulation, which was highly variable between subjects, with the variability of the BOLD activity just before the action onset. We reasoned that this latter parameter should be linked to the fluctuations of the internal state of the cortex and should, therefore, reflect the BOLD activity associated with endogenous neural rhythms. Figure 7A plots, for each subject, the relationship between BOLD variability in V1 at time 0 s against the magnitude of the SOA modulation of V1 BOLD. The correlation is high and statistically significant (Pearson's r = 0.487, p = 0.024). Furthermore, the same correlation is observed between the variability of M1 BOLD at time 0 s and the visual BOLD modulation (Pearson's r = 0.52, p = 0.015). This result demonstrates that the pre-action BOLD variability in M1 explains some of the individual differences in the modulation of the future BOLD visual response of V1 at times 6-12 s. To control for overall amplitude modulation, we replicated the same analysis using the variability of the residuals of the GLM fitting, replicating the same pattern (not shown).

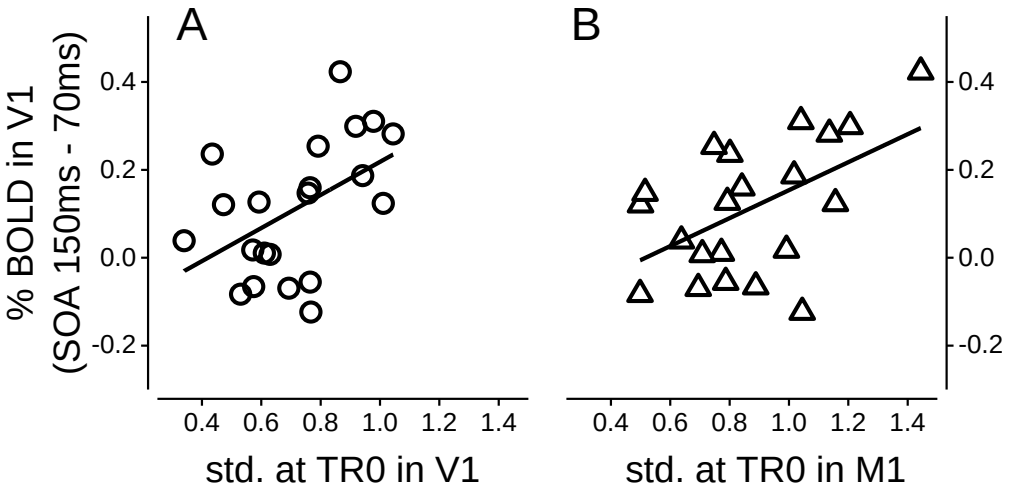

Figure 7. Magnitude of V1 BOLD oscillation correlates with BOLD variance before task. A: % BOLD difference in V1 between stimuli presented at SOA 150 ms and SOA 70 ms from button-press, as a function of BOLD variability (standard deviation, std.) in V1 at time zero from vision and action events. Each circle represents one participant. The thick line shows the linear regression (Pearson's r = 0.487, p = 0.024). **B**: same as in A, for BOLD variability in M1. Each triangle represents one participant. The thick line shows the linear regression (Pearson's r = 0.52, p = 0.015).

Model of the BOLD response component of the visuo-motor system

The overall pattern of results strongly suggests a robust temporal binding between visual and motor cortex activity during a simple visuo-motor task: a button-press and a visual discrimination of a stimulus presented at different SOAs. The pattern of results is complex with many properties not observed before: 1) the BOLD modulation is not simply multiplicative but has a stronger negative effect at later times; 2) the modulatory signal affects all V1, not only the region directly stimulated; 3) the modulatory signal is already present at time 0, given the correlation between the BOLD variance at this time with the V1 BOLD modulation with SOA; 4) the BOLD modulation is strongly correlated with perceptual accuracy; 5) the BOLD response in V1 is very different in active and passive condition; 6) the V1 is modulated by the voluntary action even in the absence of visual stimuli (see Figure 1S). To understand the main mechanisms necessary to explain simultaneously all these effects in V1, we developed a toy model simulating some well-known physiological properties of the visuo-motor loop, including signals during motor preparation phase, oscillatory rhythms and gain modulation of visual responses.

We modelled the readiness potential activity associated with the keypress with a ramp signal starting about 1500 ms before the action onset (Libet et al., 1983; Shibasaki and Hallett, 2006); the readiness potential accompanies all voluntary actions and it is also present when the subject reports the perceptual decision by keypress. Consistent with this idea, our results of supplementary Figure 1S show that the keypress action elicits a significant BOLD response in V1. Using a standard hemodynamic function, we fitted the V1 BOLD response to determine the ramp amplitude that best fitted the data (R^2 = 0.527, dark yellow trace of Figure 8B, ramp

amplitude shown in Figure 8A). The standard hemodynamic response also fits the visual response when the stimulus is observed passively ($R^2 = 0.903$, black trace in Figure 8A and 8B). The amplitude of the visual input was chosen to maximize the goodness of fit for the passive visual response. To model the evoked BOLD time-course in active visual conditions (Figure 2C), we added a negative signal synchronous with action onset, which we assume to be related to the synchronization of endogenous oscillation with the action onset. BOLD activity in the visual cortex is highly sensitive to visual noise and produces a smaller response when the components of the noise spectra are synchronized in phase (Perna et al., 2008; Rainer et al., 2001). We reasoned that a similar effect may take place with endogenous oscillations: BOLD excitability may be larger when the oscillations are in random rather than in synchronized phase. We simulate this signal as a boxcar that starts at action onset and lasts for several seconds after (Figure 8A, red trace) in agreement with previous findings (Benedetto et al., 2016; Zhang et al., 2019). The sum of these 3 independent components captures the major characteristic of the time-course in the active conditions, such as the BOLD response for 70 and 150 ms SOA. Given the strong evidence for increased cortical excitability at specific phases of theta oscillations, we postulated that visual stimuli presented at different SOAs would generate responses with different multiplicative gain. This would suggest that we can model the difference of the response at 70 or 150 ms SOA simply by scaling the amplitude of the delta function of Figure 8A (black trace). The variation in gain needed to fit the data is very small, about 33%, which would represent the amplitude that the synchronization of endogenous oscillations produces in the theta range. The fit (not shown for simplicity) is equally good for the response to the other two SOA stimulus presentations (230 and 310 ms).

To sum up, simulating some well-known biological mechanisms, such as readiness potential, sensitivity of visual cortex to noise, phase resetting of endogenous oscillations and gain change with theta oscillations, allowed us to fit 6 different time-courses (36 points) well, using only 5 free parameters.

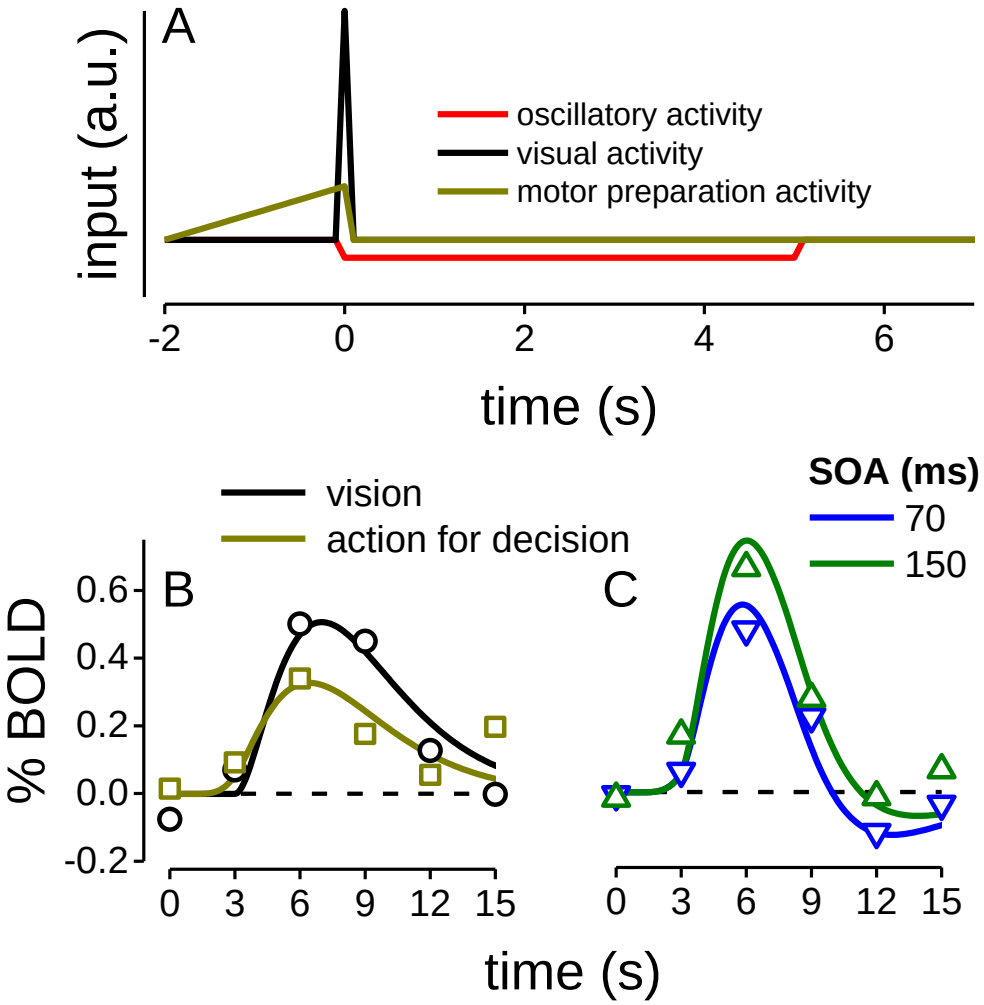

Figure 8. Model of V1 BOLD response for the different SOAs. A: Model of the biological visual and motor signals that elicit the BOLD responses. The visual signal (black line) is a delta function at the time of stimulus presentation; the motor preparation activity is modeled by a ramp starting about 2 s from button-press (dark yellow line); theta synchronization after button-press is modeled by a negative boxcar that starts at the time of action onset and lasts for several seconds afterwards (red line; the component before button-press is implicit in the ramp function). **B**: Simulated V1 BOLD response to visual events presented without keypress to start the trial obtained by convolution of the delta function in A with a standard hemodynamic function (black curve). Black circles: measured BOLD responses as in Figure 2D. Simulated V1 BOLD response to "*action for decision*" obtained by convolving the motor preparation ramp signal with the same function (dark yellow ramp in A). Dark yellow squares: measured BOLD responses as in inset of Figure S1B. **C**: Simulated V1 BOLD response obtained from the convolution with the standard hemodynamic of the sum of all signals in A when the delta amplitude (black line in A) is attenuated or amplified by 20% for the 70 ms (blue) and 150 ms (green) SOA model respectively. Hemodynamic function (from Boynton et al., 1996) with decay time $\tau = 2$ s, peak delay $\delta = 3$ s and number of cascade $n = 3$. Triangles data from Figure 3A.

DISCUSSION

By recording the BOLD response to brief transient stimuli after variable delay from action onset, we demonstrated that the BOLD response in V1 fluctuates rhythmically at about 5 Hz, following closely the rhythmic behavioral discrimination performance (Figure 1). The functional connectivity between V1 and M1 was stronger at the times of higher discrimination accuracy (Figure 6). The state of the motor cortex before the stimulus presentation correlates with the BOLD modulation at the various visuo-motor delays across the individual subjects (Figure 7). In addition, the visual response after the action is different from that recorded during passive viewing, given the small but statistically significant motor response measured in V1. All these effects could be well modelled by assuming rhythmic synchronization of V1 and M1 activity, which may be instrumental in mediating visuo-motor temporal binding.

Motor preparation and execution, such as finger tapping, button press, reaching or grasping actions and even isometric muscle contraction, can modulate the perception of visual stimuli (for a review see Benedetto et al., 2020). From hundreds of milliseconds before action onset (during motor preparation) to up to 1 second after action offset, visual sensitivity oscillates in synchrony with action onset at a frequency in the theta range (Benedetto et al., 2016; Tomassini et al., 2015; Zhang et al., 2019). Here, we replicated this behavioral oscillation using a discrimination task suitable for fMRI experiment, and consistently observed oscillations at 5 Hz. Given that some discriminations, such as contrast increment sensitivity, require the high precision coding of primary visual cortex, we reasoned that the oscillation should modulate V1 activity. In the present work we demonstrate that V1 BOLD responses follow the same rhythmic modulation at 5 Hz, strongly implicating V1 as their origin. This result is consistent with single cell recordings in monkeys showing rhythmic theta range modulation in V1 (Bosman et al., 2012; Spyropoulos et al., 2018) and with EEG recording in humans during an active task (Tomassini et al., 2017).

In the period long before the voluntary action onset (1.5 s) Theta oscillations, which are typical of motor or hippocampal activity (Fuentemilla et al., 2010; for a review see Buzsaki, 2002), predict discrimination of stimuli presented up to several hundreds of milliseconds after the execution of the movement (Tomassini et al., 2017). The prevailing model considers that the phase of the theta motor oscillations, synchronized during the preparation phase of the action, rhythmically excite V1 circuitry and hence boost BOLD responses (Tomassini et al., 2017). Whether this is a simple epiphenomenal consequence of a dense connectivity in the brain between all cortices, or has a functional role is still highly debated. We believe that our results strongly favor an important functional role of synchronous oscillation in the visuo-motor interaction or coupling, given the correlation of the effect with the variability of M1 BOLD before the stimulus appearance (Figure 7, see later in the discussion).

Many researchers have associated the synchronized neuronal oscillation to the natural fluctuations of spatial attention (VanRullen, 2016; VanRullen et al., 2007). Attention can modulate V1 and the BOLD response of V1 (Gandhi et al., 1999; for a review see Carrasco, 2011). Spatial allocation of attention to different visual positions or different objects or target features results in behavioral oscillations at different frequencies or

phases (Fiebelkorn et al., 2013; Huang et al., 2015; Jia et al., 2017; Landau and Fries, 2012; Mo et al., 2019; Re et al., 2019). It has been put forward that the oscillations reveal the dwelling time of attention that rhythmically shifts between attractors. In our experiment, the BOLD oscillation may be related to a rhythmic shift of attention between the upper and lower visual field. This might be a possible interpretation, provided that the shift in allocation of attention is in synchrony with the motor preparation. However, two results suggest this is unlikely. First, Tomassini et al. (2015) measured behavioral oscillations for the ipsilateral and contralateral positions of the stimuli with respect to the action and reported that the position did not influence systematically the behavioral oscillation. Second, we observed a small but significant BOLD modulation with stimulus time presentation also in the representation of the far peripheral visual field, which is not excited by our stimuli, and attention would not be allocated to it. Attention boosts the BOLD response at the attended stimulus position, but also increases the negative BOLD at other retinotopic representations, but we observed the same phase modulation of BOLD with SOA for central and far periphery eccentricities. In addition, the modulation has the same phase for the upper and lower visual fields, in contrast with several behavioral results showing an anti-phase relationship between different spatial locations (Fiebelkorn et al., 2013; Landau and Fries, 2012). This reinforces the suggestion that the oscillations are an expression of the sensory-motor loop, and that attention is not the primary driver.

It has been suggested that, to achieve optimal visuo-motor coordination, the brain may use a phase code of theta rhythms, as supported by data in monkeys (Fiebelkorn and Kastner, 2019). Similarly, EEG and fMRI simultaneous recording in humans (Hanslmayr et al., 2013) during contour integration tasks have demonstrated that the phase of endogenous theta oscillation predicts the functional connectivity between the visual cortical areas mediating the discrimination. Our results corroborate this evidence, in showing that the functional connectivity between V1 and M1 is stronger at SOAs corresponding to peaks of the behavioral discrimination. This is not a simple consequence of a stronger BOLD response, because it is also seen in the residual signal after removing the response to the stimulus. Nor is it driven by a general aspecific variation of the physical signal of the scanner, because it is specific to M1 and not to a closely located area, like S1.

Sensory elaboration of embodied stimuli or stimuli perceived during active scanning is different from when the stimulus is not task relevant or is passively observed (Blakemore et al., 1998). Not only does the magnitude of the peak response change, but also the temporal dynamics. A well-described phenomenon is the motor-suppression of the visual response and perception, in the first 100 ms after the action onset. The phenomenon is very obvious after saccadic eye movements (Burr et al., 1994), but it is also present after hand movements (Blakemore et al., 1998; Wolpert et al., 1995). The attenuation of the response to stimulus presentation at 70 ms delay might be interpreted as the consequence of the motor-inhibition. However, the suppression also occurs periodically at later times, questioning the concept of motor suppression (Benedetto and Morrone, 2017; Benedetto et al., 2016). It may not be a simple transient suppression, but a cyclic alternation that precedes action onset, and lasts for several seconds afterward. In this view, the function of the modulation is not to eliminate unstable visual signals collected when sensors move or attention is allocated to moving effectors but,

rather, part of a cyclic mechanism of temporal binding. *Intentional binding* signals have been repeatedly observed in monkeys' V1 cortex and may serve important functions in perception (Engel and Singer, 2001; Fries, 2015; Neupane et al., 2017). One of these functions may be related to modulation of temporal integration or temporal segmentation of the visual processing, as demonstrated in V1 during saccade in an MEG study (Wutz et al., 2016).

The strong coupling between V1 and M1 could lead to oscillatory theta modulation of the M1 BOLD. Interestingly, we observed this only in a condition where it was important to keep the discrimination decision in memory for a long time before responding about 15 s after the keypress onset. This result may suggest that visual processing, which oscillates, reaches M1 to induce oscillation there, too. Several previous papers have shown that reaction times also oscillate at theta frequencies and are reset by visual or attentive cues (Huang et al., 2015; Song et al., 2014). The stimulus-related modulation of M1 observed here may be similarly related to the programming of the decision response. However, here the task was spatial frequency discrimination, and the time of the stimulus was uninformative. Nevertheless, M1 was modulated by time. This suggests that the memory of the stimulus always contains the complete representation of it, and not only the informative variables. Memory shares much circuitry that mediates sensory elaboration. Working memory may use theta oscillations to reactivate cyclically the memory trace (Duzel et al., 2010; Fuentemilla et al., 2010), and the oscillations could mediate binding of the various sensory analyses necessary for memory recall (Lee et al., 2005). In our experiment the observed theta oscillation in M1 may derive from the cyclical recall of the memory trace of the stimulus during motor preparation phase. Whatever the mechanism generating the BOLD modulation with stimulus delay, it is important to stress that M1 activity during motor preparation can predict some of the visual BOLD modulation, implying that a common mechanism is active in the visual and motor cortex.

Interestingly, it is the variability of the M1 BOLD that is the important variable that explains some of the variance of the visual BOLD oscillatory modulation across subjects. The amplitude of endogenous oscillations is related to cortical excitability (Romei et al., 2008) and this could change at the individual level. Large variations in excitability (producing large M1 or V1 BOLD variability) should be associated with greater behavioral temporal modulation. But what could be the role of the temporal synchrony between M1 and V1? Motor cortex has been implicated in many studies of perceptual time. M1 can code temporal prediction of sensory stimuli and supplementary motor cortex code visual temporal intervals (Coull and Nobre, 1998; Protopapa et al., 2019). During action, visual and tactile time appear compressed (Haggard et al., 2002; Morrone et al., 2005; Tomassini et al., 2018), implicating, again, that motor cortex is involved in time perception. Our result, showing the predictive power of motor BOLD in explaining the oscillation effect, is consistent with the idea of a prevailing role of M1 in time perception. However, if this is associated with the functional role of a master clock located in M1, or somewhere else in the brain, it is still an open question.

None of our BOLD measures rely on the assumption of particular hemodynamic responses. This allowed us to observe that the difference in V1 response to stimuli presented in passive or active conditions develops over

time. To simulate the visual response in the active condition, we need to add a negative and long-lasting BOLD response (the red component of Figure 8). We postulated that this signal represents the endogenous oscillation that becomes synchronized to produce a prevailing theta rhythm. This is in agreement with recent evidence that the negative post-stimulus undershoot of the BOLD response is correlated with endogenous oscillations (Mullinger et al., 2013). The V1 response is usually higher to visual noise than to structured patterns of the same spatio-temporal spectra. We reasoned that the overall BOLD excitability would be higher to endogenous oscillations in random phase than when they are congruent in phase, suggesting negative BOLD after action. This is consistent with the evidence of negative BOLD in the presence of synchronized endogenous rhythms (Becker et al., 2011; Scheeringa et al., 2011a). This negative signal should last for a long time to model the changes in the BOLD response but, again, this is consistent with a large body of literature (Benedetto and Morrone, 2017).

Our results also show that motor action alone induces a V1 response. To model this effect, we considered that the preparatory motor signal (*readiness potential*; Libet et al., 1983) can modulate V1 and modify the amplitude of this gradual signal to model the BOLD response. This signal may have the important function of synchronization of V1 activity. Synchronization is crucial for efficient communication between distant areas (Canolty et al., 2006; Fries, 2015; Roelfsema et al., 1997), and it is associated with amplitude and directionality change of the BOLD responses (Becker et al., 2011; Scheeringa et al., 2011a; Scheeringa et al., 2011b). However, these two signals cannot explain the theta BOLD oscillation in V1. To this end we have to postulate that the visual response *per se* changes with cortical excitability, which should reflect theta rhythms. While the synchronous oscillation generates a simple subtracting signal, cortical excitability should change visual response in a multiplicative manner by changing the neuronal gain. By considering that response gain changes at theta frequency, we can model all the complex dynamics of the V1 response as a function of stimulus-action onset time. All the properties used to model the response are biologically plausible and already described in different contexts. However, in our task, the dissection of the individual mechanisms aids in the understanding of the phenomenon and of the role of the synchronization of the endogenous oscillation.

The effect of action on vision has been studied from the seminal intuition of corollary discharge of Helmholtz and, later, Sperry (Helmholtz, 1866; Sperry, 1950; von Holst and Mittelstaedt, 1950). However, the details of the underlying neuronal mechanisms are still largely unknown. For visual perception during eye movements it is well established that pre-motor signals can dramatically change visual processing and perception (for a review see: Binda and Morrone, 2018; Wurtz, 2018). Similar strong modulations are now apparent during locomotion in the primary cortex of mice (Niell and Stryker, 2010). These complex effects can be generalized to the complete visual-motor loop: the execution of an action always generates a change in primary visual cortex (Gallivan et al., 2019; Gutteling et al., 2015; Monaco et al., 2020; Straube et al., 2017). Straube et al. (2017) observed the suppression of the V1 BOLD response to visual stimulation caused by an action. This has been explained in the predictive code of the future consequence of the action (Brown et al., 2013; Gallivan et al., 2019; Rao and Ballard, 1999; Wolpert et al., 1995), also needed to elaborate a sense of agency. Our result

reinforces the idea of temporal oscillation as a mechanism that mediates temporal binding, which is necessary both for predictive coding and for the sense of agency: they have the temporal precision to mediate binding and the great advantage to carry prediction.

METHODS

Experimental Model and Subject Details

Human subjects

Experimental procedures are in line with the declaration of Helsinki and were approved by the regional ethics committee (Comitato Etico Pediatrico Regionale—Azienda Ospedaliero-Universitaria Meyer—Firenze (FI)) and by the Italian Ministry of Health, under the protocol 'Plasticità e multimodalità delle prime aree visive: studio in risonanza magnetica a campo ultra-alto (7T)' version #1 dated 11/11/2015. Written informed consent was obtained from each participant, which included consent to process and preserve the data and publish them in anonymous form.

We recruited seven participants for the behavioral experiment (26-36 years old, mean age = 30, 3 females and 4 males), with normal or corrected to normal vision. Twenty-four healthy participants with normal or corrected-to-normal vision were recruited for the fMRI experiment. Three of them did not complete the full protocol with a complete acquisition of stimulus response and were excluded from the analyses. The remaining 21 participants (24-63 years old, mean age = 30, 9 females and 12 males), were delivered with 84±30 visual presentations (mean ±standard deviation) during 5 fMRI acquisitions.

Method Details

Behavioral experiment: setup and procedure

Behavioral measurements were taken, in a quiet room in dim light condition. The stimuli were generated in MATLAB (MATLAB r2010a; MathWorks), displayed with a graphics board ViSaGe (Cambridge Research System) on a linearized CRT monitor (Barco Calibrator 40×30 cm, resolution of 800×600, refresh rate of 100 Hz, mean luminance 25 cd/m^2) at 57 cm from the observer. Keypress responses were acquired by CB6 response box (Cambridge Research System).

The visual stimulus comprised two horizontal gratings (10% contrast) presented for 30 ms in the upper and lower visual field, both vignetted by a Gaussian window (σ = 5 deg) at screen center (see inset Figure 1 and 2). The two gratings were separated by a Gaussian horizontal blank with space constant of 0.5 deg (σ). The spatial frequency of the gratings was 1 cpd and 1.1 cpd (10% increase) at random in the upper or lower visual field. The phase of each grating was varied randomly across trials. The subject maintained fixation on a small red dot (0.15 deg) at the center of the display, and pressed a key at will (with the fourth finger of the right hand). The visual stimulus was displayed at 60 possible stimulus onset asynchronies (SOAs) from button press, chosen randomly on each trial from a uniform distribution between 0 and 600 ms in steps of 10 ms. The subject's task was to discriminate whether the upper or the lower grating had the higher spatial frequency (i.e. a 2AFC spatial frequency discrimination task) by pressing a button with the index or the middle finger of the right hand.

Participants were instructed to wait at least 1 s between two successive keypresses (this interval was 1.6 s on average, with 1 s standard deviation). Trials that did not respect this minimum delay (1.3% of trials) were excluded from further analysis. We collected 1008±187 trials for each participant (mean and standard deviation).

Behavioral oscillation analysis

Visual accuracy as a function of SOAs was computed (as percentage of correct response) on the aggregate observer dataset, obtained by pooling all trials from all participants. Accuracy was calculated in nonoverlapping bins of 20 ms with respect to the keypress and best fitted with:

$$y(x) = \beta_0 + \beta_1 \sin(2\pi f x) + \beta_2 \cos(2\pi f x)$$

where f represents the frequency of the oscillation and is allowed to vary between 2 and 20 Hz in steps of 0.1 Hz; β_0 ,β_1 and β_2 are free parameters to be estimated. Statistical significance was evaluated by comparing the goodness of fit (R2) obtained from the original dataset, with the distribution of goodness of fits obtained on permuted datasets created by shuffling the original time stamps (10000 permutations, with replacement) with frequency f as a free parameter. This approach enables to correct for multiple frequencies comparisons.

fMRI experiment: setup and procedures

Visual stimuli were presented by using a magnetic resonance-compatible goggle set (VisuaStimDigital, Resonance Technologies Inc., Northridge, California) with visual field of approximately 32×24 deg, a resolution of 800×600 pixels, refresh rate 60 Hz, mean luminance 25 cd/m^2. Keypresses were recorded through a magnetic resonance-compatible response-box (Evoke Response Pad System, Resonance Technologies Inc., Northridge, California), positioned on the right hand of volunteers.

The same visual stimulus and procedure of the behavioral experiment were used in the fMRI scans, with minor differences. To accommodate for the different refresh rate of the monitor, the duration of the stimulus 33 ms. Participants were instructed to wait at least 15 s between two successive keypresses. As for the behavioral experiment, to start the trial, participant pressed a button on the response-box below the fourth finger; to indicate a response to either the upper or lower visual field, they pressed the button corresponding to either the index or middle finger, respectively. In addition for some trials the participant were not required to report their visual discrimination judgment by keypress and we refer to this condition as "*no action for decision*" condition to distinguish it from the condition where the subject reported the decision by keypress ("*with action for decision*" condition). In the " *no action for decision*" condition, visual stimuli were presented every 18 s on average (3 s standard deviation); in the "*with action for decision*", stimuli were presented every 30 s on average (15 s between the visual stimulus and the perceptual judgment report, and another 15 s between the perceptual judgment report and the next visual stimulus, standard deviation: 4 s). In the scan we tested four possible SOA between the visual stimulus and the keypress (70, 150, 230, or 310 ms), selected to match the maxima and minima of performance as assessed in the behavioral experiment. The average number of trials per participant

and per SOA was: 25 (sd = 10) for SOA 70 ms; 25 (sd = 11) for SOA 150 ms; 18 (sd = 6) for SOA 230 ms; 14 (sd = 8) for SOA 310 ms. For technical issues, like multiple keypress to start the trials or fast RTs, a total of 5 acquisitions (over 3 participants) were discarded from the analysis; furthermore, the 310 ms SOA was not presented to 4 subjects. In a subset of N = 5 participants, the SOAs 150 and 180 ms were additionally tested.

All participants (N = 21) were tested in the "*no action for decision*" condition; 19 of them were also tested in the "*with action for decision*" condition. A small subset of participants (N = 5) in the pilot phase of the experiment were tested in passive condition where the visual stimulus was delivered by the experimenter mimicking the inter-stimulus intervals of the active task "*with action for decision*" (mean and standard deviation interval between stimulus onset and keypresses: 16 ± 2 s). In this condition participants performed the action only to report the perceptual decision (*action for decision*).

For each participant, we performed scans to estimate retinotopic mapping of visual areas, by using meridian stimuli, defined by two 45° wedges centered around the horizontal or vertical meridian. The horizontal and vertical meridians were presented alternately for 5 TRs each (without blanks), six times for a total of 60 TRs.

MRI scanning

Scanning was performed on a Discovery MR950 7T whole body MRI system (GE Healthcare, Milwaukee, WI, USA) equipped with a 2-channel transmit coil driven in quadrature mode, a 32-channel receive coil (Nova Medical, Wilmington, MA, USA) and a high-performance gradient system (50 mT/m maximum amplitude and 200 mT/m/ms slew rate).

For twelve participants, the anatomical images were acquired at 1 mm isotropic resolution using a T1-weighted magnetization-prepared Fast Spoiled Gradient Echo (FSPGR) sequence with the following parameters: TR = 6 ms, TE = 2.2 ms. FA = 12 deg, rBW = 50 kHz, TI = 450 ms, ASSET = 2. For the remaining 9 participants, high-resolution anatomical images with at 0.8 mm isotropic resolution were acquired with a modified magnetization-prepared rapid gradient echo sequence (MP2RAGE) with the following parameters: TR = 6.6 ms, TE = 2.5 ms, FA = 5 deg, rBW= 62.5 kHz, TI = 1000 ms and 3200 ms, ARC factor = 2.

Functional images were acquired with plane spatial resolution of 1.5x1.5 mm^2 and slice thickness 1.4 mm with slice spacing = 0.1 mm, TR = 3 s, TE = 23 ms, rBW = 250 kHz, ASSET = 2, phase encoding direction: Anterior-to-Posterior. No resampling was performed during the reconstruction. For each EPI sequence, we acquired also an additional volume with the reversed phase encoding direction (Posterior-to-Anterior), used for distortion correction (see pre-processing).

Pre-processing of imaging data

Analyses were performed with Freesurfer v6.0.0, SPM12, BrainVoyager 20.6, and FSL version 5.0.10 packages.

Anatomical images were processed by a standard procedure for segmentation implemented in Freesurfer (recon-all). In addition, each hemisphere was aligned to a left/right symmetric template hemisphere (fsaverage_sym).

Functional images were corrected for subject movements, slice-time acquisition. Geometrical distortions were compensated by using EPI images with reversed phase encoding direction (Brain Voyager COPE plug-in). The preprocessed images were aligned to each participant's anatomical image using a boundary-based registration algorithm (Freesurfer "*bbergister"* algorithm) and projected to the cortical surface of each hemisphere. All analyses were conducted on data in the individual subject space, resampled to a resolution of 3x3x3 mm. For each run, we removed the first 2 TRs, then estimated the temporal linear trend and subtracted it from the fMRI time-course. Time-courses were averaged within each ROI (see below) and converted into percentage signal change by normalizing them to the mean BOLD.

Data Analysis

Definition of Regions of Interest (ROI)

In each participant, we defined three ROIs (bilateral V1, left M1 and left S1) in the 3D native space of each participant. V1 was defined using the responses to the vertical and horizontal meridians from the retinotopic mapping scans and included the cortical representation of the stimulus area, in both hemispheres of about 7.3 cm^3. The M1 and S1 ROIs were defined based on the pattern of BOLD responses to keypress associated with the perceptual decision that generated two separate foci located rostrally and caudally of the left Rolandic sulcus, consistent with M1 and Brodmann Area 3a. The M1 and S1 ROIs comprised on average 5.8 cm^3 of cortex.

Evaluation of fMRI Activity

We used the deconvolution approach for estimating the Hemodynamic Response Function (HRF) over a window of 7 TRs following each event (an "event" being a visual stimulus presentation, a voluntary keypress or both). This analysis approach does not assume a particular shape of the hemodynamic response but, rather, estimates it directly by modelling each of the 7 TRs after the events as a separate GLM predictor and thereby assigning each with a beta-weight (Costagli et al., 2014). Given that visual stimuli and keypress that started the trial were separated by less than 350 ms, we aligned the predictor with the TR that included the keypress. In each ROI, beta values at 6 TR were estimated with GLM for each type of events (e.g. for the "*no action for decision*" condition tested in all participants, there are four types of events corresponding to the 4 SOAs; for the "*with action for decision*" condition there is an additional fifth type of event corresponding to the perceptual decision) plus a constant; the average goodness of fit was 34%.

Our main analyses were performed after extracting GLM beta-weights in individual participants and in each of our *a priori* defined ROIs. When comparing values across conditions, we used linear mixed-effect model analysis (implemented in R version 4.0.2, with the function '*lme*' part of the '*nlme*' package). Full factorial

model was used with three fixed effects (TASK, SOA, and TASK-SOA interaction) and subject as random effects. Full factorial model was also adopted for fixed effect of SOA.

We assessed pairwise correlations between ROIs in the aggregate observer data, after regressing out the event-related modulations, i.e. using the residuals of the GLM model. Specifically, we performed 6 steps: 1) compute the predicted time-course given the estimated GLM weights for each run; 2) subtract it from the observed time-course to obtain the residuals; 3) concatenate residuals from all runs and all participants; 4) select one event type (one of the four SOAs); 5) average residuals within the preselected time-window 6-12s, separately for each event of the selected type and for the three ROIs; 6) compute the Pearson's correlation between residuals in the V1 and M1 ROIs, or V1 and S1 or S1 and M1. Statistical significance was assessed via bootstrapping with 10000 repetitions (with replacement), which means that step #4 was repeated 10000 times, every time selecting a different set of events.

We also generated surface maps of the GLM beta weights at the TRs of interest (the peak of hemodynamic response at TR2 for Figure 2 and Figure S1, or the average beta weights within TR2 and TR4 for Figure 4) by projecting the preprocessed BOLD timecourses from each participant on a common mirror-symmetric template of the cortical surface, concatenating them and then applying the GLM analysis. For Figure S1, data from the two hemispheres were analyzed separately; for Figures 2 and 4, the concatenated BOLD timecourses were averaged across hemispheres before the GLM analysis. Activity maps were masked for statistical significance ($p_{FDR} < 0.05$). Eccentricities in primary visual cortex were defined according to the Benson template for fsaverage (Benson et al., 2012).

BIBLIOGRAPHY

Aliu, S.O., Houde, J.F., and Nagarajan, S.S. (2009). Motor-induced suppression of the auditory cortex. J Cogn Neurosci 21, 791-802.

Becker, R., Reinacher, M., Freyer, F., Villringer, A., and Ritter, P. (2011). How ongoing neuronal oscillations account for evoked fMRI variability. J Neurosci 31, 11016-11027.

Benedetto, A., Burr, D.C., and Morrone, M.C. (2018). Perceptual Oscillation of Audiovisual Time Simultaneity. eNeuro 5, ENEURO.0047-0018.2018.

Benedetto, A., and Morrone, M.C. (2017). Saccadic Suppression Is Embedded Within Extended Oscillatory Modulation of Sensitivity. J Neurosci 37, 3661-3670.

Benedetto, A., Morrone, M.C., and Tomassini, A. (2020). The Common Rhythm of Action and Perception. J Cogn Neurosci 32, 187-200.

Benedetto, A., Spinelli, D., and Morrone, M.C. (2016). Rhythmic modulation of visual contrast discrimination triggered by action. Proc Biol Sci 283, 20160692.

Benson, N.C., Butt, O.H., Datta, R., Radoeva, P.D., Brainard, D.H., and Aguirre, G.K. (2012). The retinotopic organization of striate cortex is well predicted by surface topology. Curr Biol 22, 2081-2085.

Binda, P., and Morrone, M.C. (2018). Vision During Saccadic Eye Movements. Annu Rev Vis Sci 4, 193-213.

Blakemore, S.J., Wolpert, D.M., and Frith, C.D. (1998). Central cancellation of self-produced tickle sensation. Nat Neurosci 1, 635-640.

Bosman, C.A., Schoffelen, J.M., Brunet, N., Oostenveld, R., Bastos, A.M., Womelsdorf, T., Rubehn, B., Stieglitz, T., De Weerd, P., and Fries, P. (2012). Attentional stimulus selection through selective synchronization between monkey visual areas. Neuron 75, 875-888.

Boynton, G.M., Engel, S.A., Glover, G.H., and Heeger, D.J. (1996). Linear Systems Analysis of Functional Magnetic Resonance Imaging in Human V1. The Journal of Neuroscience 16, 4207-4221.

Brown, H., Adams, R.A., Parees, I., Edwards, M., and Friston, K. (2013). Active inference, sensory attenuation and illusions. Cogn Process 14, 411-427.

Burr, D.C., Morrone, M.C., and Ross, J. (1994). Selective suppression of the magnocellular visual pathway during saccadic eye movements. Nature 371, 511-513.

Buzsaki, G. (2002). Theta oscillations in the hippocampus. Neuron 33, 325-340.

Canolty, R.T., Edwards, E., Dalal, S.S., Soltani, M., Nagarajan, S.S., Kirsch, H.E., Berger, M.S., Barbaro, N.M., and Knight, R.T. (2006). High gamma power is phase-locked to theta oscillations in human neocortex. Science 313, 1626-1628.

Carrasco, M. (2011). Visual attention: the past 25 years. Vision Res 51, 1484-1525.

Costagli, M., Ueno, K., Sun, P., Gardner, J.L., Wan, X., Ricciardi, E., Pietrini, P., Tanaka, K., and Cheng, K. (2014). Functional Signalers of Changes in Visual Stimuli: Cortical Responses to Increments and Decrements in Motion Coherence. Cerebral Cortex 24, 110-118.

700 Coull, J.T., and Nobre, A.C. (1998). Where and when to pay attention: The neural systems for directing
701 attention to spatial locations and to time intervals as revealed by both PET and fMRI. Journal of
702 Neuroscience 18, 7426-7435.

703 Coull, J.T., Vidal, F., Nazarian, B., and Macar, F. (2004). Functional anatomy of the attentional modulation of
704 time estimation. Science 303, 1506-1508.

705 Cunnington, R., Windischberger, C., Deecke, L., and Moser, E. (2003). The preparation and readiness for
706 voluntary movement: a high-field event-related fMRI study of the Bereitschafts-BOLD response.
707 Neuroimage 20, 404-412.

708 Duhamel, J.R., Colby, C.L., and Goldberg, M.E. (1992). The updating of the representation of visual space in
709 parietal cortex by intended eye movements. Science 255, 90-92.

710 Duzel, E., Penny, W.D., and Burgess, N. (2010). Brain oscillations and memory. Curr Opin Neurobiol 20, 143-
711 149.

712 Engel, A.K., and Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness.
713 Trends Cogn Sci 5, 16-25.

714 Fiebelkorn, I.C., and Kastner, S. (2019). A Rhythmic Theory of Attention. Trends Cogn Sci 23, 87-101.

715 Fiebelkorn, I.C., Saalmann, Y.B., and Kastner, S. (2013). Rhythmic sampling within and between objects
716 despite sustained attention at a cued location. Curr Biol 23, 2553-2558.

717 Fries, P. (2015). Rhythms for Cognition: Communication through Coherence. Neuron 88, 220-235.

718 Fuentemilla, L., Penny, W.D., Cashdollar, N., Bunzeck, N., and Duzel, E. (2010). Theta-coupled periodic
719 replay in working memory. Curr Biol 20, 606-612.

720 Gallese, V., Fadiga, L., Fogassi, L., and Rizzolatti, G. (1996). Action recognition in the premotor cortex. Brain
721 119 (Pt 2), 593-609.

722 Gallivan, J.P., Chapman, C.S., Gale, D.J., Flanagan, J.R., and Culham, J.C. (2019). Selective Modulation of
723 Early Visual Cortical Activity by Movement Intention. Cereb Cortex 29, 4662-4678.

724 Gandhi, S.P., Heeger, D.J., and Boynton, G.M. (1999). Spatial attention affects brain activity in human
725 primary visual cortex. Proc Natl Acad Sci U S A 96, 3314-3319.

726 Goodale, M., and Milner, D. (2013). Sight Unseen (Oxford: OUP).

727 Goodale, M.A., Pelisson, D., and Prablanc, C. (1986). Large adjustments in visually guided reaching do not
728 depend on vision of the hand or perception of target displacement. Nature 320, 748-750.

729 Gutteling, T.P., Petridou, N., Dumoulin, S.O., Harvey, B.M., Aarnoutse, E.J., Kenemans, J.L., and Neggers, S.F.
730 (2015). Action preparation shapes processing in early visual cortex. J Neurosci 35, 6472-6480.

731 Haggard, P., Clark, S., and Kalogeras, J. (2002). Voluntary action and conscious awareness. Nat Neurosci 5,
732 382-385.

733 Hanslmayr, S., Volberg, G., Wimber, M., Dalal, S.S., and Greenlee, M.W. (2013). Prestimulus oscillatory
734 phase at 7 Hz gates cortical information flow and visual perception. Curr Biol 23, 2273-2278.

735 Helmholtz, H.v. (1866). Handbuch der physiologischen Optik (Leipzig Leopold Voss).

736 Huang, Y., Chen, L., and Luo, H. (2015). Behavioral oscillation in priming: competing perceptual predictions
737 conveyed in alternating theta-band rhythms. J Neurosci 35, 2830-2837.

738 Jia, J., Liu, L., Fang, F., and Luo, H. (2017). Sequential sampling of visual objects during sustained attention.
739 PLoS Biol 15, e2001903.

740 Kaplan, R., Doeller, C.F., Barnes, G.R., Litvak, V., Duzel, E., Bandettini, P.A., and Burgess, N. (2012).
741 Movement-related theta rhythm in humans: coordinating self-directed hippocampal learning. PLoS Biol 10,
742 e1001267.

743 Landau, A.N., and Fries, P. (2012). Attention samples stimuli rhythmically. Curr Biol 22, 1000-1004.

744 Lee, H., Simpson, G.V., Logothetis, N.K., and Rainer, G. (2005). Phase locking of single neuron activity to
745 theta oscillations during working memory in monkey extrastriate visual cortex. Neuron 45, 147-156.

746 Libet, B., Gleason, C.A., Wright, E.W., and Pearl, D.K. (1983). Time of conscious intention to act in relation to
747 onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act. Brain
748 106 (Pt 3), 623-642.

749 Melloni, L., Schwiedrzik, C.M., Rodriguez, E., and Singer, W. (2009). (Micro)Saccades, corollary activity and
750 cortical oscillations. Trends Cogn Sci 13, 239-245.

751 Merchant, H., and Yarrow, K. (2016). How the motor system both encodes and influences our sense of time.
752 Current Opinion in Behavioral Sciences 8, 22-27.

753 Mo, C., Lu, J., Wu, B., Jia, J., Luo, H., and Fang, F. (2019). Competing rhythmic neural representations of
754 orientations during concurrent attention to multiple orientation features. Nat Commun 10, 5264.

755 Monaco, S., Malfatti, G., Culham, J.C., Cattaneo, L., and Turella, L. (2020). Decoding motor imagery and
756 action planning in the early visual cortex: Overlapping but distinct neural mechanisms. Neuroimage 218,
757 116981.

758 Morrone, M.C., Ross, J., and Burr, D. (2005). Saccadic eye movements cause compression of time as well as
759 space. Nat Neurosci 8, 950-954.

760 Mullinger, K.J., Mayhew, S.D., Bagshaw, A.P., Bowtell, R., and Francis, S.T. (2013). Poststimulus undershoots
761 in cerebral blood flow and BOLD fMRI responses are modulated by poststimulus neuronal activity. Proc Natl
762 Acad Sci U S A 110, 13636-13641.

763 Neupane, S., Guitton, D., and Pack, C.C. (2017). Coherent alpha oscillations link current and future receptive
764 fields during saccades. Proc Natl Acad Sci U S A 114, E5979-E5985.

765 Niell, C.M., and Stryker, M.P. (2010). Modulation of visual responses by behavioral state in mouse visual
766 cortex. Neuron 65, 472-479.

767 Nobre, A.C., and O'Reilly, J. (2004). Time is of the essence. Trends Cogn Sci 8, 387-389.

768 Perna, A., Tosetti, M., Montanaro, D., and Morrone, M.C. (2008). BOLD response to spatial phase
769 congruency in human brain. J Vis 8, 15 11-15.

770 Protopapa, F., Hayashi, M.J., Kulashekhar, S., van der Zwaag, W., Battistella, G., Murray, M.M., Kanai, R.,
771 and Bueti, D. (2019). Chronotopic maps in human supplementary motor area. PLoS Biol 17, e3000026.

Rainer, G., Augath, M., Trinath, T., and Logothetis, N.K. (2001). Nonmonotonic noise tuning of BOLD fMRI signal to natural images in the visual cortex of the anesthetized monkey. Curr Biol 11, 846-854.

Rao, R.P., and Ballard, D.H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci 2, 79-87.

Re, D., Inbar, M., Richter, C.G., and Landau, A.N. (2019). Feature-Based Attention Samples Stimuli Rhythmically. Curr Biol 29, 693-699 e694.

Roelfsema, P.R., Engel, A.K., Konig, P., and Singer, W. (1997). Visuomotor integration is associated with zero time-lag synchronization among cortical areas. Nature 385, 157-161.

Romei, V., Brodbeck, V., Michel, C., Amedi, A., Pascual-Leone, A., and Thut, G. (2008). Spontaneous fluctuations in posterior alpha-band EEG activity reflect variability in excitability of human visual areas. Cereb Cortex 18, 2010-2018.

Ross, J., Morrone, M.C., Goldberg, M.E., and Burr, D.C. (2001). Changes in visual perception at the time of saccades. Trends in Neurosciences 24, 113-121.

Scheeringa, R., Fries, P., Petersson, K.M., Oostenveld, R., Grothe, I., Norris, D.G., Hagoort, P., and Bastiaansen, M.C. (2011a). Neuronal dynamics underlying high- and low-frequency EEG oscillations contribute independently to the human BOLD signal. Neuron 69, 572-583.

Scheeringa, R., Mazaheri, A., Bojak, I., Norris, D.G., and Kleinschmidt, A. (2011b). Modulation of visually evoked cortical FMRI responses by phase of ongoing occipital alpha oscillations. J Neurosci 31, 3813-3820.

Schroeder, C.E., Wilson, D.A., Radman, T., Scharfman, H., and Lakatos, P. (2010). Dynamics of Active Sensing and perceptual selection. Curr Opin Neurobiol 20, 172-176.

Shibasaki, H., and Hallett, M. (2006). What is the Bereitschaftspotential? Clin Neurophysiol 117, 2341-2356.

Song, K., Meng, M., Chen, L., Zhou, K., and Luo, H. (2014). Behavioral oscillations in attention: rhythmic alpha pulses mediated through theta band. J Neurosci 34, 4837-4844.

Sperry, R.W. (1950). Neural basis of the spontaneous optokinetic response produced by visual inversion. J Comp Physiol Psychol 43, 482-489.

Spyropoulos, G., Bosman, C.A., and Fries, P. (2018). A theta rhythm in macaque visual cortex and its attentional modulation. Proc Natl Acad Sci U S A 115, E5614-E5623.

Stetson, C., Cui, X., Montague, P.R., and Eagleman, D.M. (2006). Motor-sensory recalibration leads to an illusory reversal of action and sensation. Neuron 51, 651-659.

Straube, B., van Kemenade, B.M., Arikan, B.E., Fiehler, K., Leube, D.T., Harris, L.R., and Kircher, T. (2017). Predicting the Multisensory Consequences of One's Own Action: BOLD Suppression in Auditory and Visual Cortices. PLoS One 12, e0169131.

Tomassini, A., Ambrogioni, L., Medendorp, W.P., and Maris, E. (2017). Theta oscillations locked to intended actions rhythmically modulate perception. Elife 6.

Tomassini, A., Spinelli, D., Jacono, M., Sandini, G., and Morrone, M.C. (2015). Rhythmic oscillations of visual contrast sensitivity synchronized with action. J Neurosci 35, 7019-7029.

808 Tomassini, A., Vercillo, T., Torricelli, F., and Morrone, M.C. (2018). Rhythmic motor behaviour influences
809 perception of visual time. Proc Biol Sci 285, 20181597.

810 VanRullen, R. (2016). Perceptual Cycles. Trends Cogn Sci 20, 723-735.

811 VanRullen, R., Carlson, T., and Cavanagh, P. (2007). The blinking spotlight of attention. Proc Natl Acad Sci U
812 S A 104, 19204-19209.

813 von Holst, E., and Mittelstaedt, H. (1950). Das Reafferenzprinzip. Naturwissenschaften 37, 464-476.

814 Wolpert, D.M., Ghahramani, Z., and Jordan, M.I. (1995). An internal model for sensorimotor integration.
815 Science 269, 1880-1882.

816 Wurtz, R.H. (2018). Corollary Discharge Contributions to Perceptual Continuity Across Saccades. Annu Rev
817 Vis Sci 4, 215-237.

818 Wutz, A., Muschter, E., van Koningsbruggen, M.G., Weisz, N., and Melcher, D. (2016). Temporal Integration
819 Windows in Neural Processing and Perception Aligned to Saccadic Eye Movements. Curr Biol 26, 1659-
820 1668.

821 Yokosaka, T., Kuroki, S., Nishida, S., and Watanabe, J. (2015). Apparent time interval of visual stimuli is
822 compressed during fast hand movement. PLoS One 10, e0124901.

823 Zhang, H., Morrone, M.C., and Alais, D. (2019). Behavioural oscillations in visual orientation discrimination
824 reveal distinct modulation rates for both sensitivity and response bias. Sci Rep 9, 1115.

825
826

Tables

Table 1. Effects of SOA and task in V1, M1 and S1

	df		V1 F	V1 p-val	V1 par. η^2	V1 Model Comp.*	M1 F	M1 p-val	M1 par. η^2	M1 Model Comp.*	S1 F	S1 p-val	S1 par. η^2	S1 Model Comp.*
			Y ~ SOA both conditions											
SOA	3	56	4.57	**0.006**	0.197	t_{20}=3.16 p=**0.004**	0.90	0.443	0.046	t_{20}=1.02 p=0.317	3.09	**0.034**	0.142	t_{20}=1.94 p=0.06
			Y ~ SOA*TASK											
SOA	3	124	3.74	**0.012**	0.083		2.69	**0.048**	0.061		6.85	**0.0003**	0.142	
TASK	1	124	0.86	0.353	0.007		0.38	0.536	0.003		0.70	0.402	0.006	
SOAxTASK	3	124	0.50	0.680	0.012		5.18	**0.002**	0.112		5.14	**0.002**	0.111	
			Y ~ SOA for significant SOAxTASK interaction with action for decision											
SOA	3	50					3.87	**0.014**	0.189	t_{18}=3.22 p=**0.004**	3.79	**0.015**	0.185	t_{18}=3.12 p=**0.005**
			Y ~ SOA no action for decision											
SOA	3	56					1.00	0.396	0.051	t_{20}=-0.81 p=0.425	2.61	0.060	0.123	t_{20}=0.42 p=0.674

SOA: *70, 150, 230, 310 ms*; **TASK**: *with action for decision, no action for decision*

*****: *Paired t-test: 70∪230 ms vs 150∪310 ms*