A Non-negative Measure Of Feature-specific Information Transfer Between Neural Signals

Jan Bím[1,5], Vito De Feo[1], Daniel Chicharro[1, 2], Ileana L. Hanganu-Opatz[4],

Andrea Brovelli[3 # *], Stefano Panzeri[1 # *]

1. Center for Neuroscience and Cognitive Systems@UniTn, Istituto Italiano di Tecnologia, Rovereto (TN) 38068, Italy.
2. Department of Neurobiology, Harvard Medical School, Boston, MA 02115, USA
3. Institut de Neurosciences de la Timone, UMR 7289, Aix Marseille Université, CNRS, 13385 Marseille, France
4. Developmental Neurophysiology, Institute of Neuroanatomy, University Medical Center, Hamburg-Eppendorf, 20251 Hamburg, Germany
5. Department of Computer Science, Czech Technical University in Prague, 120 00 Prague, Czech Republic

* Corresponding authors:
Stefano Panzeri (stefano.panzeri@iit.it)
Andrea Brovelli (andrea.brovelli@univ-amu.fr)
# Co-senior authors with equal contribution

Abstract

Quantifying the amount and content of information transfer between neural populations is crucial to understand brain dynamics and cognitive functions. Most data-driven methods exploit the notion of Wiener-Granger causality, a statistical concept based on temporal prediction. Transfer Entropy and Directed Information formalise this notion by means of information theoretical quantities and can capture any (linear and nonlinear) time-lagged conditional dependencies, thus quantifying the amount of information flow between neural signals. Nevertheless, none of these metrics can reveal what type of information is exchanged. To address this issue, we developed a new measure called Feature-specific Information Transfer (FIT) that is able to quantify both the amount and content of information transfer between neuronal signals. We tested the novel metric on simulated data and showed that it successfully captures feature-specific information transfer in different communication scenarios including feedforward communication, external confounding inputs and synergistic interactions. Moreover, the FIT measure displayed sensitivity to modulations in temporal parameters of information transfer and signal-to-noise ratios, and correctly inferred the directionality of transfer between signals. We then tested FIT's ability to track feature-specific information flow from neurophysiological data. First, we analysed human electroencephalographic data acquired during a face detection task and confirmed current hypotheses suggesting that information about the presence of an eye in a face image flows from the contralateral to the ipsilateral hemisphere with respect to the position of the eye. Second, we analysed multi-unit activity data recorded from thalamus and cortex of rat's brain, and showed that the FIT measure successfully detected bottom-up information transfer about visual or somatosensory stimuli in the corresponding neural pathway. Third, we analysed cortical high-gamma activity estimated from human magnetoencephalographic data during visuomotor mapping, and confirmed the notion that visuomotor-related information flows from superior parietal to premotor areas. Altogether our work suggests that the FIT measure has the potential to uncover previously hidden feature-specific information transfer from neural data and provide a better understanding of brain communication.

Introduction

Cognitive functions arise from the dynamic coordination of neural activity over large-scale networks (Bressler and Menon, 2010; Varela et al., 2001; von der Malsburg et al., 2010). Tracking how neural populations and brain regions interact and exchange information is therefore crucial for modern neuroscience. To quantify the degree of interaction, functional connectivity (FC) measures include various forms of statistical dependencies between neural signals, such as linear correlation and Granger causality measures (Bressler and Seth, 2011; Brovelli et al., 2004; Ding et al., 2006; Granger, 1980; Seth et al., 2015). Transfer Entropy (Schreiber, 2000) and Directed Information (Massey, 1990) represent the most general model-free measures of Wiener-Granger causality, and capture any (linear and nonlinear) time-lagged conditional dependence between neural signals (Vicente et al., 2011; Wibral et al., 2014).

Directed Information, however, does not cast light on the information content – i.e., what type of information is carried. Ince and collaborators have proposed a conditional (i.e., stimulus-specific) measure of Directed Information (DI), called Directed Feature Information (DFI) (Ince et al., 2015). However, due to its theoretical construction, the DFI can be negative. This poses questions regarding its suitability as a measure of feature-specific information transfer. In the current paper, we exploit a recent extension of Shannon Information theory, called Partial Information Decomposition (Williams and Beer, 2010) to develop a novel non-negative measure of feature-specific information transfer. The Partial Information Decomposition (PID) framework allows to quantify the information that several sources provide individually (*unique* information), shared (*redundant* information) or jointly (*synergistic* information) about a target variable. Most importantly, these terms are non-negative. Within this framework, we propose a novel metric named Feature-specific Information Transfer (FIT), which formalizes this notion as a form of information redundancy between the past of the sender and the present activity of the receiving node, unique with respect to the past of the receiving node. Here, we investigated the relevance of the proposed approach on simulated data formalizing different scenarios of information transfer between two signals. Then, we tested FIT's ability to track feature-specific information flow from neurophysiological data recorded from human participants in cognitive settings (EEG and MEG) and from rats (multi-unit activity) during multisensory stimulations.

Results

Feature-specific Information Transfer (FIT)

Feature-specific Information Transfer (FIT) is a model-free measure that quantifies the amount of information transfer from a neural signal X (i.e., the sender) to a neural signal Y (i.e., the receiver) *about* an experimental variable S. Such *feature* may represent any experimental variable under investigation, such as stimulus type, behavioral response or experimental condition. The formulation of the FIT exploits latest advances in the field of information theory, in particular the Partial Information Decomposition framework (Williams and Beer, 2010). Within this framework, FIT is defined as the *redundant* information about S shared between the present of a receiving node Y_t and the past of the sending node X_{t-d}, which is *unique* (i.e., absent) with respect to the past of the receiver Y_{t-d} (see Materials and Methods for details about the derivation of the FIT measure).

We performed numerical simulations capturing multiple scenarios of stimulus-specific information transfer to test the significance of the FIT measure. We designed three scenarios exemplifying prototypical patterns of information transfer that could occur between neural populations. The first scenario modeled a feedforward stimulus-specific transfer from X to Y (Fig. 1A left). As expected, both DI, DFI and FIT captured the transfer and showed a significant peak at 50 msec (Fig. 1A). However, the DFI displayed negative values in the baseline period compared to the FIT which, by definition, is non-negative. The biased-corrected versions of the information measures showed a similar behavior, as shown in the bottom row in (Fig. 1A). The second scenario mimicked a transfer of information that cancelled in Y (Fig. 1B left). The aim of this scenario was to test the ability of different measures to distinguish between stimulus-dependent and stimulus-independent information transfer. Since the DI lacks the ability to discern stimulus-dependent from stimulus-independent information transfer, as expected, we observed a significant increase (Fig. 1B, left panel). However, both the DFI and FIT correctly reflected the lack of stimulus-specific transfer (Fig. 1B right panels). In the third scenario (Fig. 1C), we simulated a condition in which an external confounding signal was injected to the receiving node Y, without passing through the communication channel between X and Y (Fig. 1C, left panel). The DI, both bias-corrected and uncorrected, erroneously indicated an information transfer between X and Y. However, neither the DFI nor FIT measures showed a significant peak. We should note, however, that the DFI, as in Fig. 1B, exhibited significant

negative values aligned with the transfer. Overall, these results indicated that the FIT measure successfully tracked feature-specific information transfer and, unlike the DFI, it provided non-negative estimates.

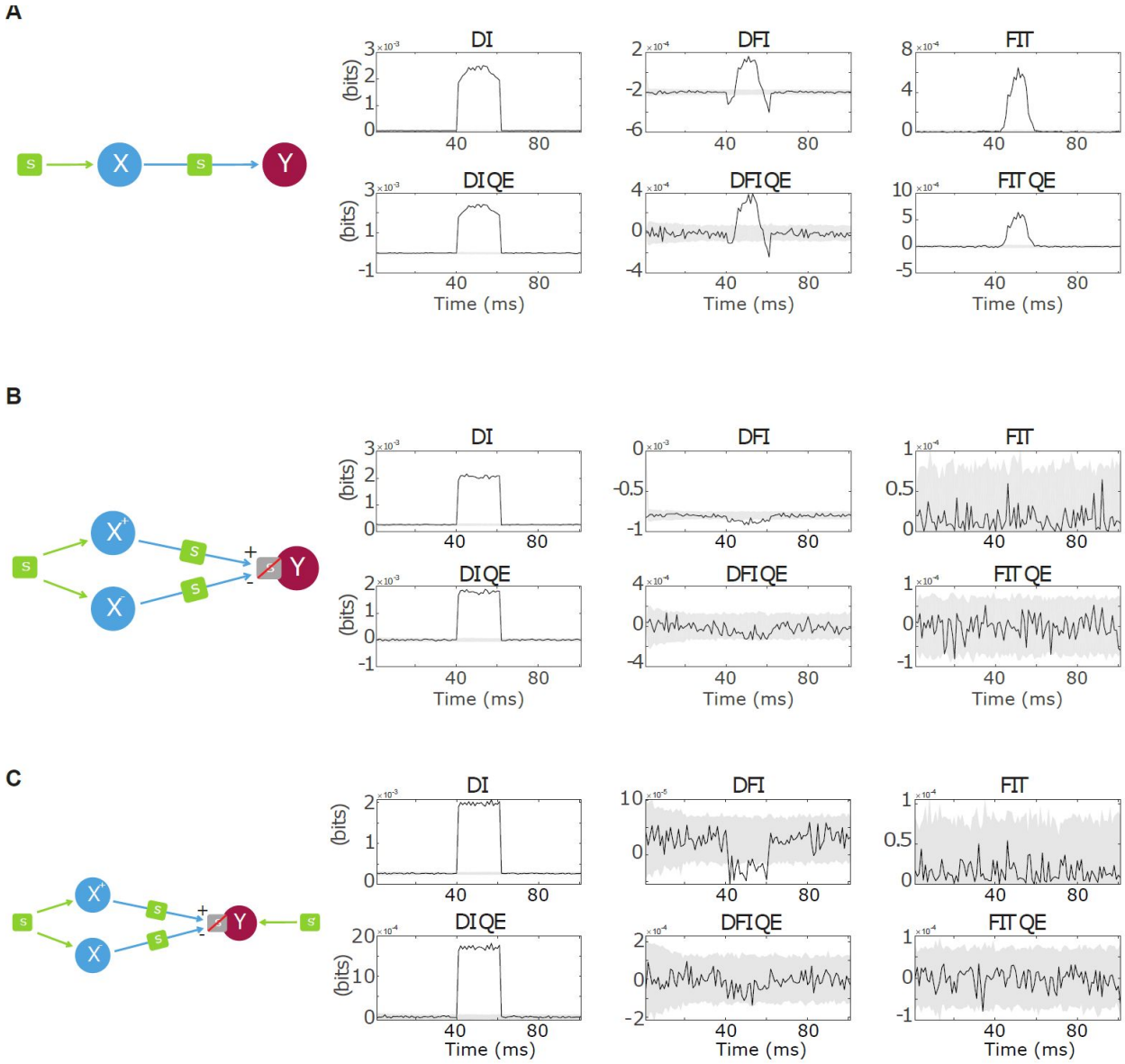

Figure 1. Three communication scenarios testing for Feature-specific Information Transfer. A Scheme of a simulation scenario representing the most basic transfer of information about S from X to Y without any extra confounders (left panel). On the right, all information quantities clearly show a significant peak where the communication occurs in their both uncorrected and bias corrected (QE) versions. It is noteworthy though, that peaks of DFI are much smaller in comparison to its surrogates (grey area). **B.** Scheme of a simulation scenario representing an information transfer from X to Y that does not convey information about S, though the information is present in X. In the right panel, first, DI, both corrected and uncorrected, correctly reflects an information transfer between X and Y. On the contrary, neither the DFI nor FIT show a significant peak. **C**. Scheme of a simulation scenario representing an information transfer from X to Y that does not convey information about S, though the information is present in X. And an extra source of information about a different stimulus S' drawn from the same probability distribution as S (left

panel). Whereas the DI shows a increase in information transfer, both the DFI and FIT correctly detect the absence of stimulus-specific information transfer. We should note, however, that the DFI contains significant negative values.

In order to demonstrate that FIT reflects the amount of stimulus information transfer, rather than a global information transfer between neural signals, we performed further numerical simulations using the first scenario by varying the noise level (i.e., α parameter) and stimulus amplitudes (i.e., β parameter) in the signals. The results showed that DI increases both with noise and stimulus level (Fig. 2, left panel). As expected, DI does not dissociate between noise and amplitude components of the information transfer. On the other hand, the FIT measure decreased with increasing noise level (Fig. 2A mid-top panel) and increases with stimulus gain (Fig. 2B mid-bottom panel). Such behavior is further exemplified in Fig. 2 (right panels), and thus demonstrates the feature-specificity of the FIT measure.

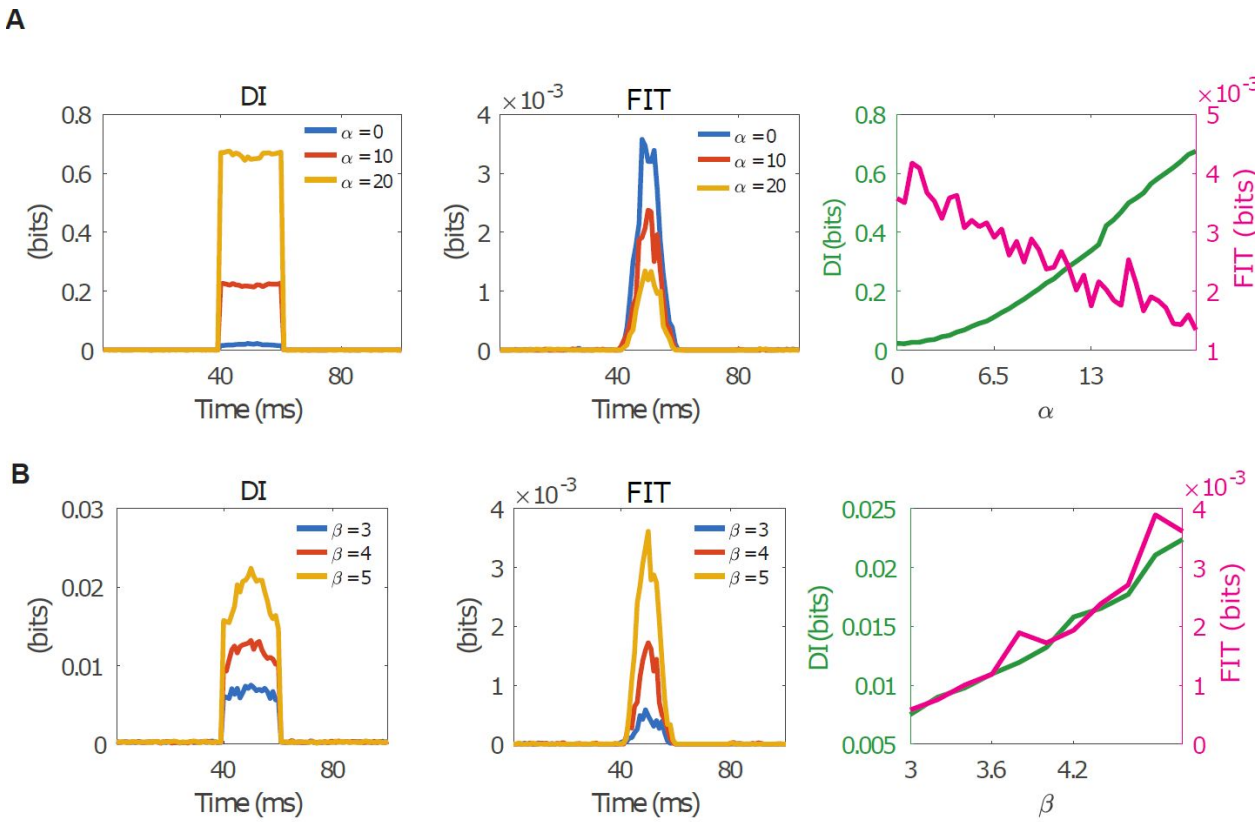

Figure 2. Dependence on noise level and signal amplitude. A. Dependence on noise level (α coefficient) on DI (left panel) and FIT (central panel). Whereas the DI increase with noise level, the FIT metric shows a decreasing pattern (central and right panels). **B**. Dependence on stimulus amplitude (β coefficient) on the DI (left panel) and FTI (central panel). Both the DI and FIT increase with stimulus amplitude (right panel).

In the brain, information transmission between neural populations may display large variability in duration and occur at variable delays depending on the neural pathway and experimental condition under investigation. We thus performed numerical simulations to study the influence of key temporal parameters of information flow on the DI, DFI and FIT measures. To do so, we performed simulations using the first scenario (Fig. 1A) and varied: i) the width of the information peaks in signals X and Y (40 and 80 msec); ii) the temporal interval between peaks (5 and 20 msec); iii) delays with respect to the peak in Y at which the information measures were computed (10, 30 and 50 msec prior to the information peak in Y). Please note that in all simulations, information transfer peaked at 60 msec. The results showed a sensitivity of the DI and DFI with respect to the width of the information peak: the time course of the DI and DFI was narrower for short width values (third row in Fig. 3B) than for long width values (first and second rows in Fig. 3B). However, the DI and DFI measure did not reflect modulations of either the temporal interval between peaks (first versus second row in Fig. 3B) or the delay at which the measures were computed (different colors in Fig. 3B). On the other hand, the FTI measure reflected modulation in the delay between the past X and present of Y (different colors in Fig. 3B right panels). For the shortest delay (10 msec, red curves in Fig. 3B), stimulus-specific information dropped earlier than for longer delays (50 msec, purple curves in Fig. 3B, right panels). The dependence on delays was absent for peak widths of 80 msec and peak intervals of 5 and 20 msec (first and second rows of Fig. 3B right panels, respectively). However, a dependence on the temporal interval between peaks was observed - i.e., the time courses of the FIT for short width values (third row in Fig. 3B) was narrower than for long width values (first and second rows in Fig. 3B).

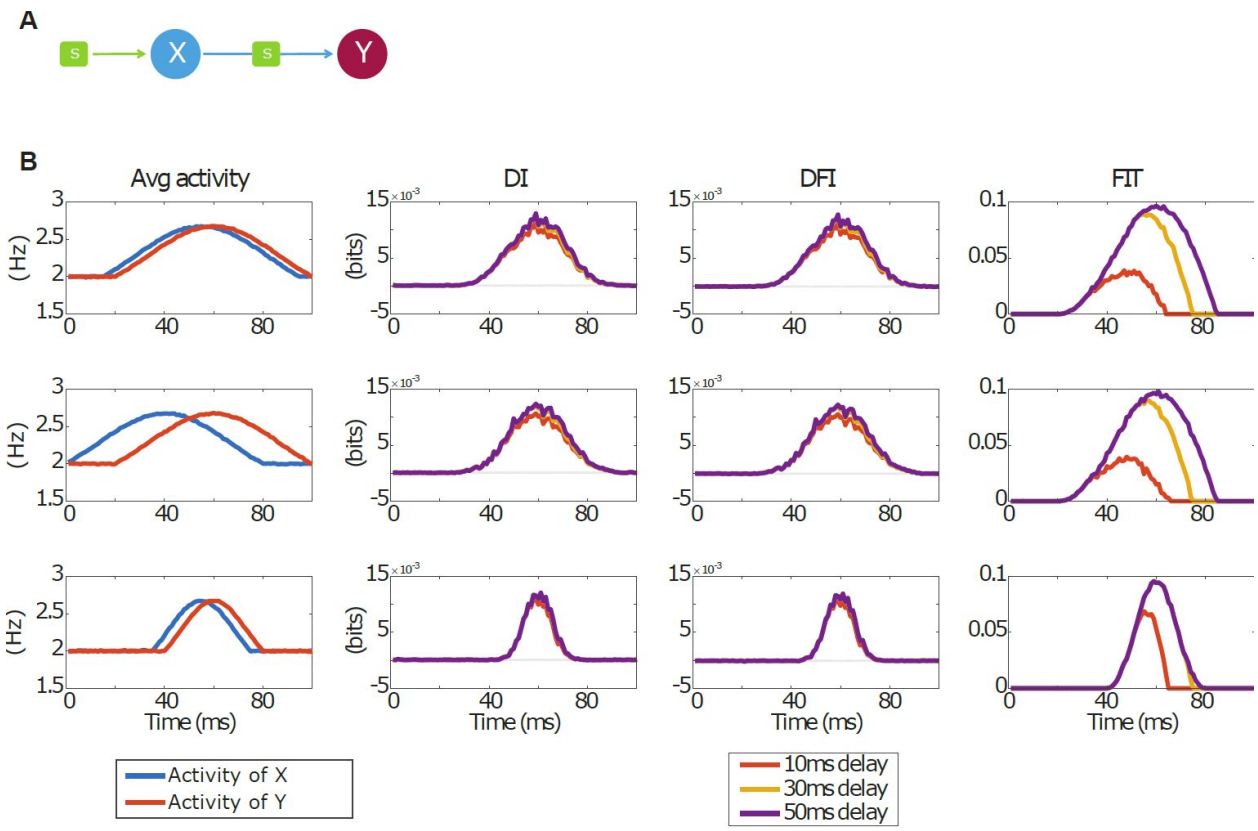

Figure 3. Dependence on temporal parameters. Simulations using the basic transfer scenario **A.** The first row in **B** depicts simulation results for wide information peaks in X and Y (80 msec) and short interval between peaks (5 msec); the second row shows simulations with wide information peaks in X and Y (80 msec) and long interval between peaks (20 msec); the third row reflects narrow peaks (40 msec) and short interval between peaks (5 msec). The left column shows the mean activity for X (blue curve) and Y (red curve). The different colors in the 2nd to 4th columns are associated with different delays at which the information measures were computed (10, 30 and 50 msec prior to the information peak in Y).

We then addressed a key question in neural data analysis which pertains the detection of directionality of information transfer between neural populations. To investigate the ability of the DI, DFI and FIT measures to detect correct directionality, we performed simulations to test directionality sensitivity based on the first scenario with signal peaks width of 80 msec and distances between peaks of 5, 15 and 25 msec (different rows in Fig. 4). Figure 4 shows that the DI and DFI did not provide clear directional information when the delay was short with respect to the overlap (top row in Fig. 4). However, the FIT clearly demonstrated consistent behavior across all parameters, correctly capturing the correct direction of transfer of information from X to Y (Fig. 4, right panels).

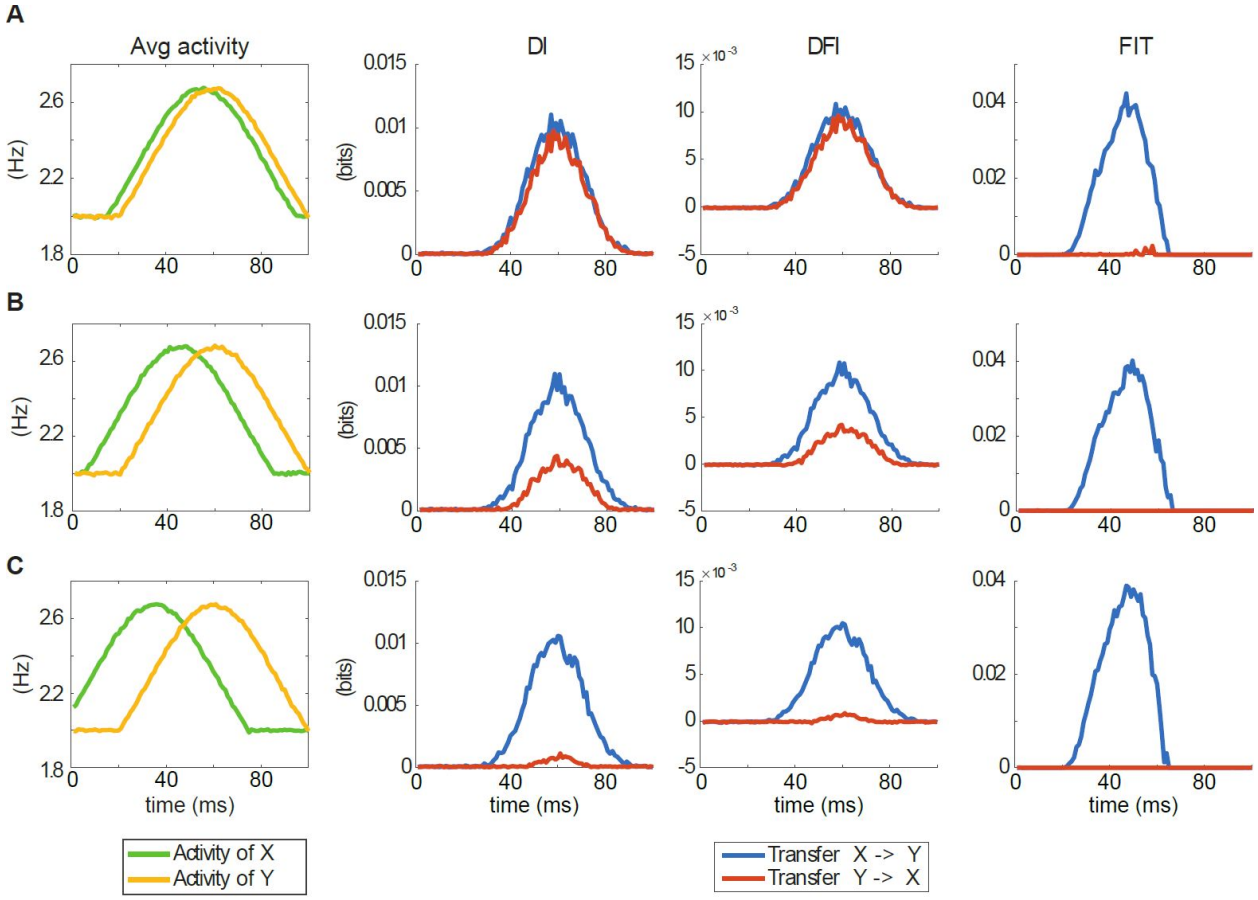

Figure 4. Directionality detection. Information flow analysis from X to Y (blue curves) and from Y to X (red curve).The left column shows the mean activity of X (green curve) and Y (yellow curve). The DI, DFI and FTI measures are displayed in the second, third and fourth columns, respectively. **A.** Short distances between peaks (5 msec) and large overlap of neural activity. **B.** Intermediate distance (15 msec). **C.** Long distance (25 msec).

To further instigate the robustness of the FIT to noise, we performed additional simulations using the first scenario and tested noise levels at the limit of detectability for the DI and DFI measures (Suppl. Fig. 1A). In the high-noise condition, we added a sine wave with a period of 200 msec and random phase at each trial to both signals X and Y. The low-noise condition had a constant value added to both signals, thus decreasing the relative difference between the baseline and the peak. Indeed, the results showed that, whereas the DI and DFI could not capture the information transfer significantly from the data (Suppl. Fig. 1B, second and third columns), the FIT was more robust to noise (Suppl. Fig. 1B, fourth column).

Then, we investigated how the different measures behave when synergistic effects are present. Indeed, the mathematical decomposition of the DFI in different terms of the PID framework (Williams & Beer, 2010) shows that the DFI contains several synergistic terms, which

may lead to negative values of the measure (please refer to the Supplementary Methods for the mathematical proof). We thus designed a scenario (Suppl. Fig. 1C) in which the co-occurrence of signal similarity and correlations of opposite sign was injected in the nodes (Pola et al., 2003). In such scenario, the stimulus-dependent correlations had opposite sign with respect to signal similarity and they were orders of magnitude higher than stimulus independent correlations. As expected, the DFI yielded significant negative values, while both DI and FIT displayed a significant increase (Suppl. Fig. 1C). Therefore, we can conclude that the FIT successfully recovered stimulus-specific information transfer even in cases of strong synergistic effects.

As a last control analysis, we investigated the susceptibility of the different measures to the sampling bias, which is known to affect information theoretical measures (Panzeri et al., 2007; Panzeri and Treves, 1996). We tested the behavior of the DI, DFI and FIT on all three scenarios with varying number of samples (Suppl. Fig. 2). The results confirmed that FIT values grow quadratically with logarithm of number of samples (Suppl. Fig. 2). This bias can, however, be corrected using the quadratic extrapolation procedure introduced in (Strong et al., 1998). Using the correction, we showed that it is possible to obtain reliable values of FIT with approximately a 10th of the number of samples.

Tracking feature-specific information flow from neurophysiological data

We analyzed three datasets to investigate the ability of the FIT measure to track feature-specific information flow from neurophysiological data.

Information transfer during face detection (human EEG)

Previous work on human face detection has shown that the rising part of the EEG response (i.e., the N170 component in the contralateral electrodes, roughly 120-170 ms post-stimulus) predominantly reflects the encoding of a single feature: the contralateral eye (Rousselet et al., 2014). More recent results (Ince et al., 2016) suggested that information about presence of an eye (left or right eye) on a picture is crucial for correct categorization. Eye-related information was suggested to flow from the contralateral to the ipsilateral hemisphere with respect to the position of the eye (Fig. 5A). Here, we tested whether the FIT measure could detect eye-specific information flow between hemispheres by means of analyses of an EEG dataset collected from human participants performing a face detection task (Rousselet et al., 2014). Subjects were asked to detect the presence of a face or a random texture while the image was covered by a

bubble mask. The FIT values clearly showed a transfer of information from the right to the left hemisphere about the presence of the left eye in the image (Fig. 5B). The increase in FIT spanned a time interval between approximately 150 to 190 msec after stimulus-onset (Fig. 5B, solid lines) and transfer delay of approximately 10 to 50 msec (Fig. 5B and D). Furthermore, we compared all four possible of transfer directions and positions of the eye on the image (i.e., from ipsi to contralateral and vice versa, and left and right eye, respectively). We confirmed that information transfer was significant exclusively from contralateral to ipsilateral hemisphere (Fig. 5C and F). A clear difference in FIT was observed between the expected directional flows: from right to left hemispheres for the left eye and from left to right hemisphere for the right eye (Fig. 5C). For the condition “left eye” in which a right-to-left transfer was observed, we compared the values in the cluster that captured the transfer (referred to as “comm”) to a cluster of the same shape and size, but positioned elsewhere on the timeline with no overlap (“no comm”). In addition, we computed the information measures for surrogate values obtained by randomly shuffling stimulus values (left-eye visibility). The results showed that FIT values were significantly higher in the communication period than those obtained at different time points and with respect to surrogates; this was true for both QE corrected and uncorrected FIT (Fig. 5E). Finally, we tried to quantify the transfer using DI and DFI and, unlike the FIT, these measures proved to be unable to capture the stimulus-specific transfer (Fig. 5F).

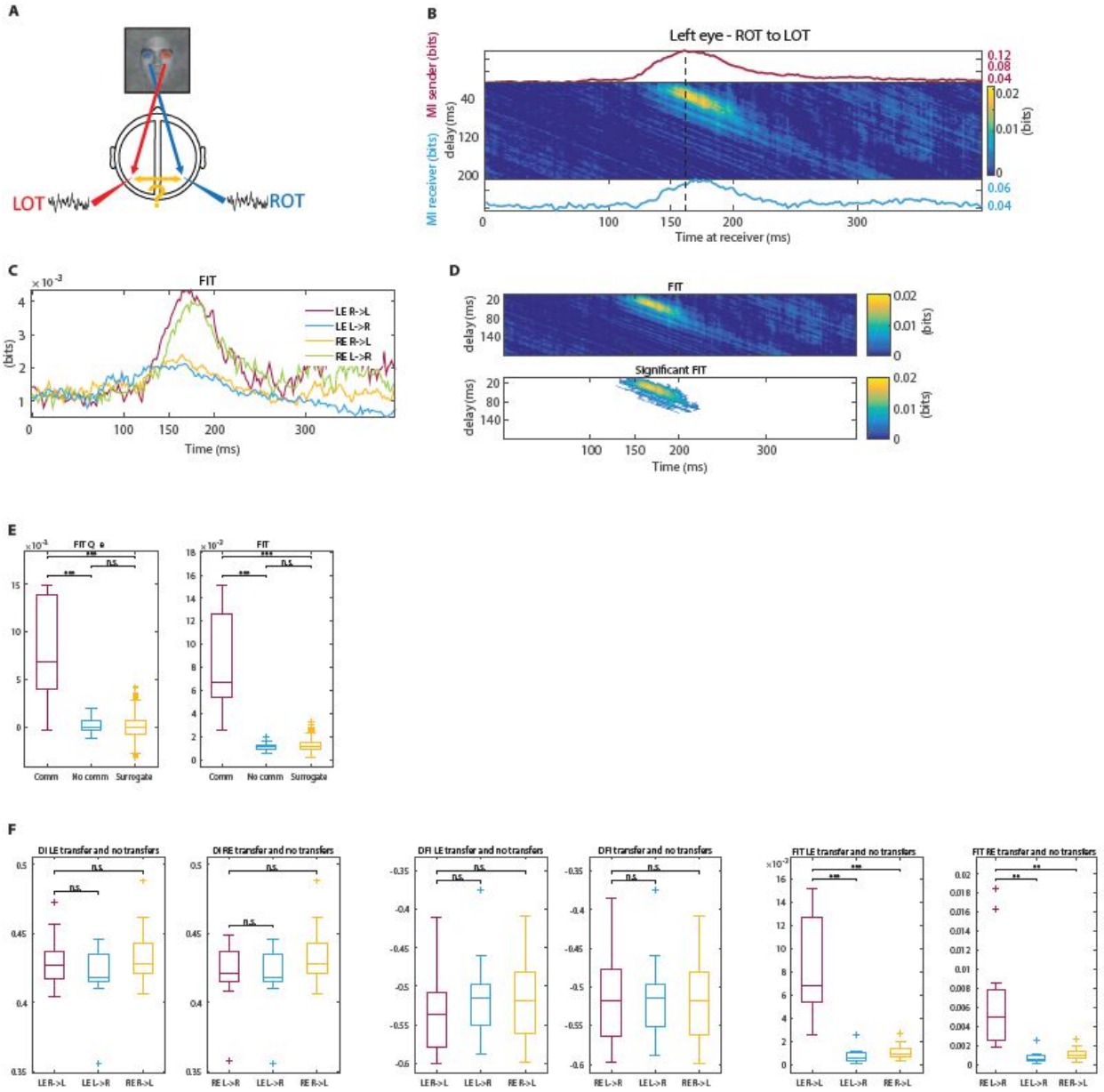

Figure 5. FIT and face detection. A. Expected information flow patterns. LOT and ROT refer to the Left and Right Occipito-Temporal sensors, respectively. **B.** FIT map (delays versus time) capturing the transfer of information about left eye from the right to the left hemisphere. Line plots above and below show mutual information between the given occipito-temporal sensor (either sending or receiving one) and visibility of the left eye. **C.** FIT values averaged across delays for all combinations of transfer direction and eye. **D** Nonparametric statistical test (Maris and Oostenveld, 2007) were used to identify significant clusters of FIT values. The upper plot shows the raw values of FIT and the bottom one displays the cluster that was identified as significant. **E.** FIT values during the transfer period (comm) compared to a cluster of the same shape and size, but positioned elsewhere on the timeline with no overlap (no comm) and surrogate values obtained by random shuffling of the stimulus values (left eye visibility). **F.** Comparison between a given contra to ipsilateral transfer (L – left, R - right) and the two ipsi to contralateral transfers for DI, DFI and FIT and both eyes (LE and RE).

Information transfer during multisensory stimulation (rodent MUA)

In the visual system, sensory information flows from the eye receptors to the primary visual cortex (V1) through the Lateral Geniculate Nucleus (LGN) of the thalamus; in somatosensory systems, information flows from the whiskers receptors to the primary somatosensory cortex (S1) through the Ventral Posteromedial nucleus (VPM) of the thalamus (Fig. 6A). The aim of our analyses was to detect the presence of stimulus-specific information transfer from thalamus to cortex. In particular, we investigated the ability of our novel measure to characterize the properties of neural information flows in two neural circuits: 1) the flow of visual information along the pathway from the eyes to the primary visual cortex (V1) through the Lateral Geniculate Nucleus (LGN) of the thalamus; 2) the flow of somatosensory information from the whiskers receptors to the primary somatosensory cortex (S1) through the Ventral Posteromedial nucleus (VPM) of the thalamus. We computed information quantities for all combinations of direction of transfer (from LGN to V1, from V1 to LGN, from SPM to VPM and from VPM to SPM). Two stimulation protocols (conditions) were analysed: 1) unimodal visual stimulation versus bimodal stimulation (simultaneous visual and tactile); 2) unimodal tactile stimulation versus bimodal stimulation (simultaneous visual and tactile). We considered only stimulation contralateral to the recording regions. We expected to observe significant FIT values for the somatosensory pathway (from VPM to S1) in condition 1, and for the visual pathway (from LGN to V1) in condition 2. Information theoretical measures were computed for every subject and then we averaged across rats.

In the somatosensory domain, the DI from the thalamus (VPM) to the cortex (S1) increased for both stimulation conditions (Fig. 6B, left and right panels). This phenomenon was observed also in the visual pathway (Suppl. Fig. 3B). For the somatosensory pathway, the DI augmented in a time interval ranging between 20 and 30 msec after stimulation (Fig. 6B). For the visual pathway, the DI augmented in a time interval ranging between 50 and 70 msec after the stimulation with a transfer lag of approximately 10 to 20 msec (Suppl. Fig. 3B). Modulations in DI were similar ($p=0.68$) across stimulation protocol (Fig. 6F, top-left panel). On the other hand, the FIT displayed a significant increase only for the stimulus feature encoded in the neural pathway (Fig. 6D and Suppl. Fig. 3D, top panels). For the somatosensory pathway, the FIT increased in a time interval from approximately 20 to 30 msec after the stimulation (Fig. 6C). For the visual pathway, the FIT increased in a time interval from approximately 50 to 70 msec after stimulation (Suppl. Fig. 3C). We then confirmed, using a non-parametric statistical test (Maris and Oostenveld, 2007), the presence of significant clusters of information transfer (Fig. 6D;

Suppl. Fig. 3D, lower panels). For the somatosensory pathway, the FIT values clearly showed a transfer of information about the presence of the tactile stimulus (Fig. 6D, right panels) and not about the presence of the visual stimulus (Fig. 6D, left panels). For the visual pathway, the FIT values clearly showed a transfer of information about the visual feature (Suppl. Fig. 3D, left panels) and not about the tactile input (Suppl. Fig. 3D, right panels). We then compared the values in the cluster that captured the transfer (comm) to a cluster of the same shape and size, but positioned elsewhere on the timeline with no overlap (no comm) and surrogate values obtained by random shuffling of the stimulus values. For the somatosensory pathway, the results showed that FIT values were significantly higher in the communication period than those obtained at different time points (p=0.0050) and with respect to surrogates (p=0.0067) (Fig. 6E, right panel); this was true also for QE corrected FIT (p=0.0031 and p=0.0039) (Fig. 6E, left panel). Also for the visual pathway, the results showed that FIT values were significantly higher in the communication period than those obtained at different time points (p=0.0045) and with respect to surrogates (p=0.0048) (Suppl. Fig. 3E, right panel); this was true also for QE corrected FIT (p=0.0034 and p=0.0048) (Suppl. Fig. 3E, left panel). Then, we compared all four combinations of information directionality (from thalamus to cortex and vice-versa) and stimulation conditions. We confirmed that significant FIT from thalamus to cortex occurred during the first 100 msec after the stimulation (p=0.032 for VPM->S1, p=0.34 for S1->VPM, p=0.019 for LGN->V1, p=0.61 for V1->LGN) (Fig. 6F and Suppl. Fig. 3F, bottom left panels).

Finally, we tested the different measures in their ability to track the correct directionality of information flow. The DI was not able to discern between stimulation conditions neither in the visual (p=0.70 for LGN->V1, p=0.67 for V1->LGN) or somatosensory domain (p=0.68 for VPM->S1, p=0.052 for S1->VPM) (Fig. 6F and Suppl. Fig. 3F, two top-left panels). The DFI, unlike the FIT, was able to capture the transfer only for the somatosensory pathway (p=0.0011 for VPM->S1, p=0.16 for S1->VPM) (Fig. 6F, top-right panels), but not for the visual pathway (p=0.125 for LGN->V1, p=0.321 for V1->LGN) (Suppl. Fig. 3F, top-right panels). The uncorrect and bias-corrected FIT provided the strongest detectability signals (p=0.046 for VPM->S1, p=0.42 for S1->VPM, p=0.043 for LGN->V1, p=0.72 for V1->LGN) (Fig. 6F and Suppl. Fig. 3F, bottom-right panels).

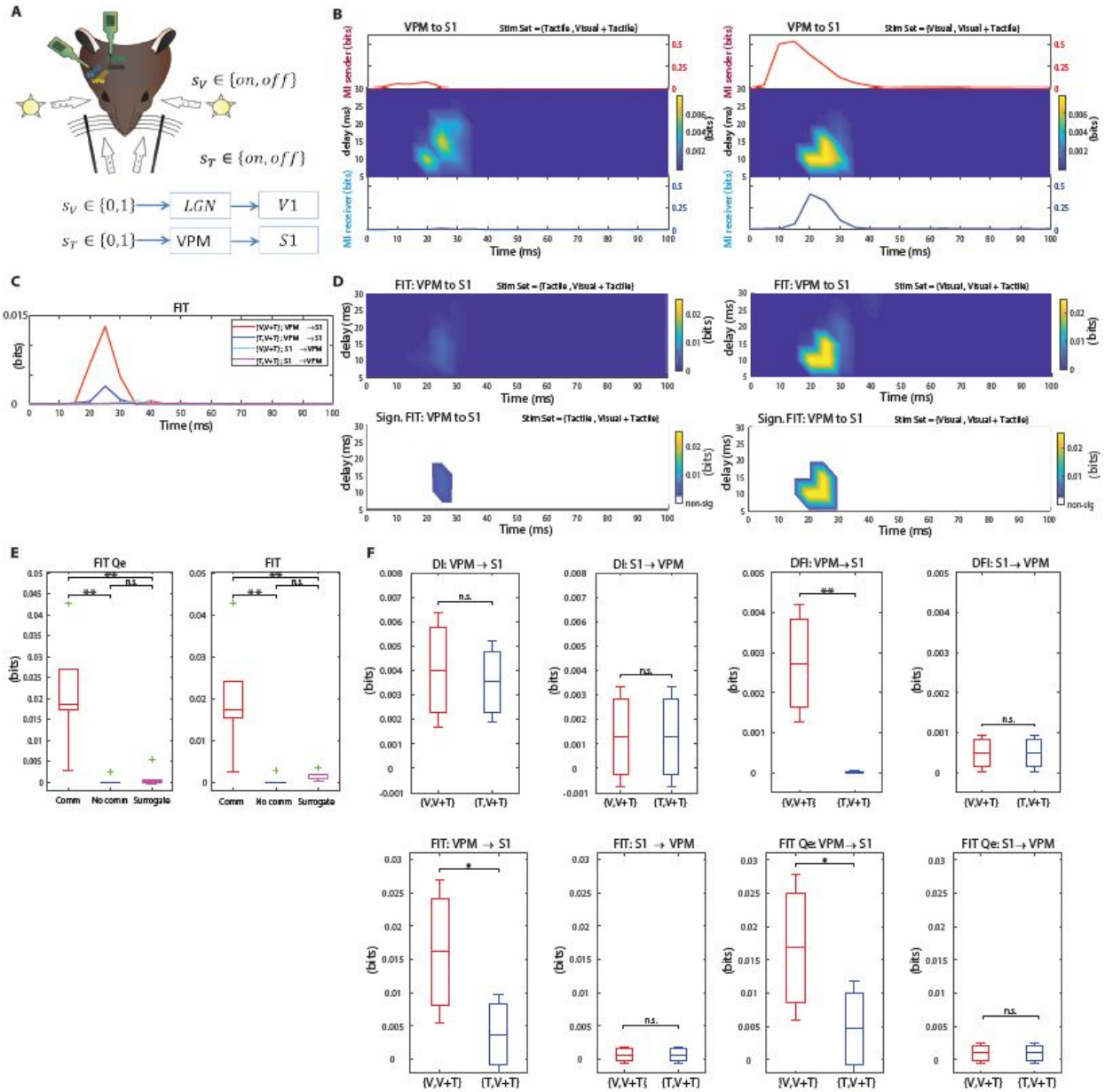

Figure 6. Information transfer in visual and somatosensory system. A. MUA was recorded from the visual and somatosensory territories of the thalamus and associated primary cortical regions. We considered only stimulation contralateral to the recording regions. Visual information flows from the eye receptors to the primary visual cortex (V1) through the Lateral Geniculate Nucleus (LGN) of the thalamus and somatosensory information flows from the whiskers receptors to the primary somatosensory cortex (S1) through the Ventral Posteromedial nucleus (VPM) of the thalamus. **B.** DI from the VPM to S1. The 2D map shows values for a given pair of time and delay that determines how far in the past the past values of X and Y were taken from. Line plots above and below show mutual information between the electrophysiological response (MUA) and the presented stimulus. **C.** FIT values on the same timeline as in B averaged across all delays for all combinations of transfer direction and stimulation. **D.** Nonparametric cluster-based statistical test (Maris and Oostenveld 2007) that identifies all significant clusters of values in the 2D map. The upper plot shows the raw values of FIT and the bottom one only the cluster that was identified as significant. **E.** Comparison of FIT value in the cluster that captured the transfer (comm) to a cluster of the same shape and size, but positioned elsewhere on the timeline with no overlap (no comm) and surrogate values obtained by random shuffling of the stimulus values. **F.** Comparisons between all 4 combinations of transfer direction (from

thalamus to cortex and vice-versa) and conditions (when the stimulus feature matched the neural pathway under examination and when it was not encoded by the brain pathway), for all the 4 measures (DI, DFI, FIT, Qe-FIT).

Information transfer in human visuomotor network (MEG)

In the third set of analyses, we used our novel metric to infer visuomotor-related changes in information transfer over the human visuomotor network. Previous work has shown that the medial superior parietal cortex (SPCm) plays a key role in the visuomotor-related FC network (Brovelli et al., 2017). Briefly, two measures of centrality, such as the eigenvector centrality and the betweenness centrality defined as the propensity of a brain area to participate in multiple sub-networks, placed the SPCm at the top of the cortical hierarchy. In particular, the SPCm was found to participate both in a visuo-parietal network early during the processing of visual information and during the planning of visuomotor associations with the dorsolateral premotor cortex (PMdl). Granger causality analyses also revealed a directional influence from the superior parietal lobe towards the premotor area (Brovelli et al., 2015).

The goal of the MEG analysis was to assess the dynamics of visuomotor-related modulations in DI, DFI and FIT as they evolve over time. The DFI and FIT were thus computed by taking as feature the event-related *versus* baseline HGA. The DI, DFI and FIT were computed between the SPCm and the dorsolateral premotor cortex PMdl. Figure 7A (left panels) shows the increase in high-gamma activity (HGA) aligned on movement onset. The DI and DFI showed a directional pattern that is consistent with previous results (higher directionality from the SPCm towards the PMdl), but the degree of asymmetry was weaker than the one observed with the FIT. In fact, the FIT was practically absent in the PMdl to SPCm direction and showed a clear drop after action execution (i.e., time equal to zero), as shown in (Fig. 7A) (center and right panels). This result supports current knowledge about the dynamics of interaction between the superior parietal lobe and the dorsolateral premotor cortex, which is expected to occur over a timescale of tens to hundreds on milliseconds during a time interval prior to motor output. Then, we analyzed the coupling between the SPCm and a companion area in the parietal lobe, the dorsal Inferior Parietal Cortex (IPCd). This analysis was performed to test the efficiency of the FIT in estimating efficiently directional coupling from neural signals from nearby brain regions that may share information and display common-driving effects. The FIT displayed a clear selectivity for one of the directionality of information flow, with respect to the DI and DFI (Fig. 7B). Overall, these results suggest that the use of the FIT to event-related analyses successfully retrieves information transfer at a high temporal and directional selectivity, key elements for tracking cortico-cortical directional interactions.

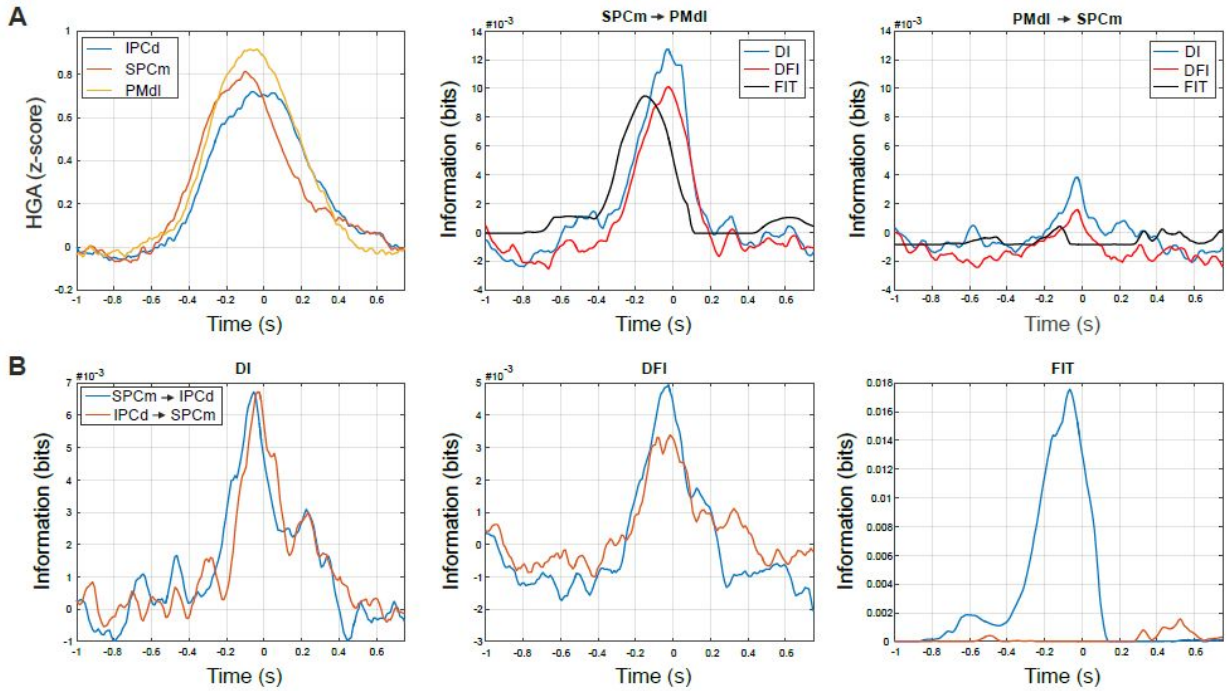

Figure 7. Event-related FIT. A. Left panel: time course of high-gamma activity (HGA) for three selected brain regions: the dorsal Inferior Parietal Cortex (IPCd), the medial superior parietal cortex (SPCm)and the dorsolateral premotor cortex (PMdl) . Central and right panels: directional influences (DI, DFI and FIT) between the SPCm and PMdl. **B.** Information flow between the SPCm and the dorsal Inferior Parietal Cortex (IPCd) for the DI (left panel), DFI (central panel) and FIT (right panel).

Discussion

In the current paper, we developed a novel information theoretic measure that quantifies feature-specific information transfer (FIT) between neural signals. The measure is based on the Partial Information Decomposition formalism (Williams and Beer, 2010) and on the notion of redundancy of information. The rationale is that transfer of information about an experimental variable, or feature, S flows between two neural signals X to Y if information about S is shared (redundant) between the past of X and present of Y, while being absent in the past of Y. Our measure therefore reflects both the amount and content of information transfer between neural signals. We performed numerical simulations to test the reliability of the FIT measure to capture feature-specific information transfer in different communication scenarios. The goal was to compare the FIT measure with two complementary metrics that are used to quantify information transfer: 1) Directed Information (DI) or Transfer Entropy, which quantifies the amount of information transfer between signals, but not its content; 2) the Directed Feature Information (DFI), which is a measure previously developed to quantify stimulus-dependent information transfer. The results of the first simulation (Fig. 1A) showed that the FIT is effective in recovering the underlying information transfer in a simple feedforward communication channel, similarly to the DI and DFI. In scenario 2, we investigated a communication channel in which the sender node contained information about the feature of interest, but did not communicate it to the receiver (Fig. 1B). The FIT successfully showed a non-significant information transfer, like the DFI, whereas the DI was positive for the whole time interval of communication. In the third scenario, which included an external confounding signal added to the receiving node (Fig. 1C), the FIT accurately represented the lack of stimulus-specific information transfer, whereas the DFI showed significant decreases in information with respect to baseline (Fig. 1C), potentially due to synergistic effects. Indeed, the DFI can be mathematically decomposed in different terms of the PID framework, which contains several synergistic (negative) terms, which may lead to negative values of the measure (Suppl. Methods). Numerical simulations using communication scenario with strong synergistic effects showed that the DFI yielded significant negative values, while the FIT displayed a significant increases (Suppl. Fig. 2C) . Thus, both theoretical and numerical results suggested that the DFI may not be taken as a metric for feature-specific information transfer in cases when synergistic effects dominate.

We then studied how the DI and FIT measures scale as a function of varying levels of noise and signal amplitude. We showed that the DI does not dissociate between noise and amplitude

components of the information transfer. On the other hand, the FIT measure decreased with increasing noise level and increased with stimulus amplitude (Fig. 2), thus suggesting that the FIT is an appropriate metric for varying signal-to-noise ratios. Subsequently, the analysis of dependence on temporal parameters of information transmission (such as the width of the information peaks in signals X and Y, the temporal interval between peaks and delays at which the information measures were computed) showed that the FIT measure correctly differentiated modulations in these parameters, whereas the DI and DFI were largely unaffected (Fig. 3). These results further confirmed the sensitivity of our measures to key temporal features of information transfer. Finally, numerical simulations testing the ability to detect the correct directionality of information transfer between neural signals showed that the FIT is more sensitive than the DI and DFI (Fig. 4). Indeed, the DI and DFI did not provide clear directional information especially when the delay was short with respect to the overlap (top row in Fig. 4), whereas the FIT clearly demonstrated consistent behavior across all parameters, correctly capturing the correct direction of transfer of information from X and Y (Fig. 4, right panels). In terms of sample bias problem, the FIT suffers the same limitation as all other Information Theory quantities and can be corrected by using Quadratic Extrapolation procedure (Panzeri et al., 2007; Strong et al., 1998) . However, simulations suggested that FIT provides information about the content of a transfer with either less data (approximately a 10th of the data size) or noisier (Suppl. Fig. 3). Overall, the analysis of numerical simulations in different communication scenarios, signal-to-noise relations, temporal parameters of information flow and directionality patterns showed that the FIT measure accurately detects feature-specific information transfer and provides a novel non-negative information theoretical measure.

A potential limitation of the FIT measure resides in the assumption that I_{min} correctly quantifies unique, redundant and synergistic information (see Materials and Methods for the definition of I_{min}). Indeed, I_{min} has been criticized for potentially miscategorising purely unique information as synergistic and there were several attempts to improve it (Bertschinger et al., 2014; Griffith and Koch, 2014; Harder et al., 2013). In the case of only two predictors, the measure of Bertschinger (Bertschinger et al., 2014) is currently preferred. However, it cannot be applied to cases when the number of predictors is larger than two, as in our case. On the other hand, we should point out that our simulations suggest that I_{min} is adequate for the communication scenarios investigated in the current study and the numerical manipulations we have performed. In the case of Gaussian variables, an alternative measure of redundancy that

provides non-negative terms is the minimum mutual information (MMI) PID (Barrett, 2015). Further work is, however, needed to assess the potential application of MMI PID in the development of a features-specific information transfer metric.

In the second part of our work, we investigated the validity of the FIT measure on neurophysiological data. The analysis of human electroencephalographic (EEG) data confirmed current hypotheses (Ince et al., 2016) suggesting that information about the presence of an eye in a face image flows from the contralateral to the ipsilateral hemisphere with respect to the position of the eye (Fig. 5). Crucially, eye-specific information transfer was detectable only by means of our novel measure, whereas neither the DI nor DFI measures successfully captured such transfer (Fig. 5F). The analysis of MUA data recorded from thalamus and cortex of rat's brain, confirmed that the FIT measure is able to detect information transfer about a specific feature (e.g. visual or somatosensory) in the correct neural pathway (Fig. 6 and Suppl. Fig. 4). FIT displayed significant increase along the somatosensory pathway from the VPM to S1 in experimental protocols that involved whiskers' stimulations (Fig. 6). In parallel, it displayed significant modulations along the visual pathway from LGN to V1 in protocols involving visual stimulations (Suppl. Fig. 4). Neither DI nor DFI could resolve feature-specific information transfer along these pathways. In addition, only the FIT metric could differentiate the correct (i.e, from thalamus to cortex) directionality of information transfer both in the visual and somatosensory domains. Finally, the analysis of MEG data confirmed previous work based on Granger causality analyses of high-gamma activity showing that the superior parietal area plays a driving role in the network, exerting Granger causality on the dorsal premotor area (Brovelli et al., 2015). However, previous work did quantify the dynamics of information flow from superior parietal regions to premotor areas. Here, we exploited our novel measure to quantify the directionality of information flow in a dynamical (time-resolved). The results showed that the FIT outperformed the DI and DFI in detecting the timing of information transfer between the superior parietal region and the dorsal premotor cortex. Indeed, the increase in FIT from the SPCm to the PMdl (Fig. 7A, central panel) peaked approximately 0.2s before the motor response, therefore confirming the notion that the dorsal frontoparietal network is crucial for the visuomotor computations transforming visual information into motor plans (Corbetta and Shulman, 2002; Culham and Valyear, 2006; Wise et al., 1996; Wise and Murray, 2000). The DI and DFI, however, were most probably influenced by the global increase in HGA observed around the motor response and their temporal dynamics decreases around motor output (Fig. 7A, the right and central panel). A second advantage of the FIT was its ability to better detect directionality of

information transfer than the DI and DFI, as demonstrated in simulations (Fig. 4). Whereas the DI and DFI show bidirectional effect, the FIT measure displayed unidirectional effect from the superior parietal to premotor are (Fig. 7B). The results suggest that the FIT reduces potential bias present in the DI and DFI due to the autocorrelation in the HGA. Taken together, the analysis of neurophysiological data suggests high selectivity for temporal and directional information and confirms the ability of the FIT to infer the dynamics of feature-specific information flow from neurophysiological data.

To conclude, the results from numerical simulations and the analysis of human and rodent neurophysiological data showed that the FIT provides high discriminability of information transfer both in time and across directions of information flow. Altogether our work suggests that the FIT measure has potential to uncover previously hidden feature-specific information transfer from neural data and provide a better understanding of brain communication.

Materials and Methods

Derivation of Feature-specific Information Transfer

In the following section, we will begin by defining concepts from Information Theory that form the cornerstones of our work. Then, we will introduce the framework of Partial Information Decomposition (PID) and the derivation of our novel measure named Feature-specific Information Transfer (FIT). In classical Information Theory, the Shannon entropy (Shannon, 1948) of the random variable X is defined as

$$H(X) = -\sum_{x \in X} p(x) \log p(x) \tag{1}$$

$H(X)$ is the expected value of outcome's *surprisal,* or *self-information,* and it defines the amount of average surprise generated by the event x. $H(X)$ represents the amount of information conveyed by X (the self-information is zero if X is fully predictable; non-negative otherwise). In addition, $H(X)$ can be interpreted as the amount of uncertainty associated with the values taken by the random variable X. The conditional entropy of X given a second variable Y is defined as

$$H(X|Y) = -\sum_{y \in Y} p(y) \sum_{x \in X} p(x|y) \log p(x|y) \tag{2}$$

$H(X|Y)$ quantifies the amount of information carried by X given that the value of another random variable Y is known. The total amount of information provided by X and Y (or the total uncertainty when observing X and Y together) is quantified by the joint entropy $H(X,Y)$:

$$H(X,Y) = H(X|Y) + H(Y) \tag{3}$$

Finally, to complete the set of all possible information interactions between two variables, the mutual information $I(X;Y)$ between X and Y is:

$$I(X;Y) = H(X) + H(Y) - H(X,Y) \tag{4}$$

All these quantities are non-negative and additive, and create a full decomposition of information content between two random variables.

The amount of transfer of information between two random processes X and Y, is usually quantified with the measures called Transfer Entropy (Schreiber, 2000) and Directed

Information (Massey, 1990), which are intimately related (Amblard and Michel, 2011). In our work, we will use the term Directed Information (DI), for the estimate

$$DI_{X \to Y} = I(Y_t \,;\, X_{t-d} \mid Y_{t-d}) \tag{5}$$

where Y_t is the value of Y at time t, whereas X_{t-d} and Y_{t-d} are the values of X and Y at some preceding point in time $t-d$ (where *d* stands for delay), respectively. DI quantifies the amount of information transferred from the past of X to the present of Y, which is not contained in the past of Y. DI is a model-free measure of Granger causality and it quantifies the amount of information transfer between two random processes.

The aim of our work was to develop a novel information theoretical metric that could quantify the amount of information transfer from X to Y *about* an experimental variable S (i.e., the content of information transfer). Such experimental *feature* may represent any experimental (dependent or independent) variable under investigation (e.g., stimulus type, experimental condition, etc.). We reasoned that three conditions must be met in order for feature-specific information transfer to occur between two random processes. First, the sender X should convey information about S at some time point in the past. Second, the receiver Y should encode such information later in time. Third, the information carried by the past of X and the present of Y should be redundant and absent in the past of Y (otherwise the transfer would not be the origin of the information). To summarize, in order for transfer of information about S to occur from X to Y, the following predicates should be satisfied:

1. The past of X at time $t-d$ should carry information about S
2. The present of Y at time t should carry information about S
3. The past of X and present of Y should share information about S that is not already present in the past of Y

The first predicate is satisfied if the mutual information between S and the past of X, namely $I(S;X_{t-d})$, is positive. Similarly, the mutual information between S and the present of Y should be non-zero. The third predicate, however, remains challenging.

A recent measure called Directed Feature Information (DFI) was introduced to quantify feature-specific directed information (Ince et al., 2015):

$$DFI^{S}_{X \to Y} = DI_{X \to Y} - DI_{X \to Y} | S \tag{6}$$

The DFI expresses the difference between the total information transfer between X and Y ($DI_{X\to Y}$) and the information transfer when the stimulus S is known ($DI_{X\to Y}|S$). The first term quantifies the total amount information transferred between X and Y , whereas the second term quantifies the amount of information that is transferred for a fixed S. Hence, the subtraction was considered as the remaining information transfer between X and Y about S. In other words, the DFI quantifies the amount of new information transferred from X to Y that is related to the variations in S (Ince et al., 2015). The DFI can also be expressed as the co-information (Bell, 2003; McGill, 1954) between X and Y with respect to S (see Supplementary Methods for the full derivation):

$$DFI^{S}_{X\to Y} = Red_S\,(Y_t\,;\,X_{t-d}\,|\,Y_{t-d}) = I\,(S;\,X_{t-d}\,|\,Y_{t-d}) + I\,(S;\,Y_t\,|\,Y_{t-d}) - I\,(S;\,X_{t-d}, Y_t\,|\,Y_{t-d}) \quad (7)$$

$Red_S\,(Y_t\,;\,X_{t-d}\,|\,Y_{t-d})$ represents the so called redundancy (Pola et al., 2003; Schneidman et al., 2003) between X_{t-d} and Y_t about stimulus S conditional on Y_{t-d}. In general, positive redundancy measures the overlap of the MI about the stimulus feature S that is present in two different responses. Negative values of redundancy indicate synergy, which occurs when the two variables carry more MI when considered together than alone (i.e., in cases when the stimulus modulates their relationship) (Ince et al., 2015). This occurs whenever the third term is larger than the information carried by the two variables observed independently. Despite the fact that DFI was successfully applied to detect feature-specific transfer (Ince et al., 2015), its property of potentially being negative hinders its interpretation as the amount of information transfer.

To clarify this issue, let us consider all information interactions between two variables X and Y (Fig. 8A). The total amount of information in the system is quantified by the joint entropy $H(X,Y)$, whereas the amount of information carried by each variable separately is quantified by the entropy values $H(X)$ and $X(Y)$, respectively. The shared part of information is quantified by the mutual information between those variables $I(X;Y)$. In order to quantify information about a single variable S, we can replace the entropies with mutual information terms with respect to S (Fig. 8B). Now, the total information about S is quantified by the mutual information $I(X,Y;S)$ and, consistently, the information separately carried by each variable is given by $I(X;S)$ and $I(Y;S)$, respectively. The shared part of information between the variables is quantified by the

redundancy between X and Y or $Red_S(X;Y)$ (i.e., corresponding to the unconditional version of eq. 7)

$$Red_S(X;Y) = I(X;S) + I(Y;S) - I(X,Y;S) \tag{8}$$

Eq. 8 demonstrated that $Red_S(X;Y)$ can be negative whenever the total amount of information about S is larger than the sum of both individual contributions, despite the fact that the shared part of information is included in the sum twice. In such case, the emergent information carried by the interaction between the two variables is called *synergistic* information. Hence, we conclude that the DFI does not satisfy the third predicate, because it does not exclusively represent information that is shared between X and Y.

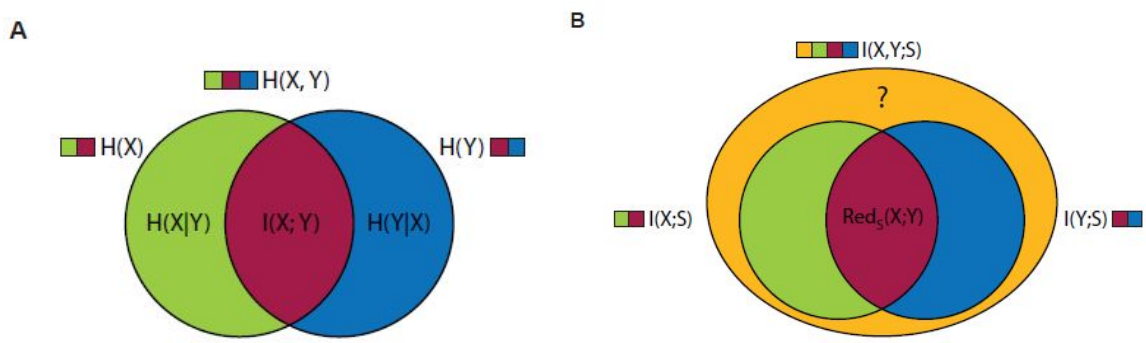

Figure 8. **A.** Information decomposition of interaction of two variables. H stands for entropy and I for mutual information. The decomposition is composed of three distinct parts, two representing information contributed by the variables independently (conditional entropies) (green and blue) and one that represent information that is provided by both of them (mutual - also called redundant or shared - information) (red). Based on the diagram it is possible to deduce that in order to compute the mutual information it is possible to add information contributed by both variables H(X) and H(Y) (1 green, 1 blue, 2 red parts) and subtract the total information H(X, Y) (1 green, 1 blue and 1 red part). **B.** Information decomposition of interaction of two mutual informations of two variables (X and Y) with a third one (S). The decomposition follows similar logic as the decomposition in (A). However, if the logic is applied to the computation of the redundant (mutual/shared) part (red) – sum of both mutual informations I(X; S) (green and red) and I(Y; S) (blue and red) and subtracting the total mutual information I(X,Y; S) – the redundant information is not non-negative. Therefore, there must be another part of the diagram that contains the extra information (yellow) that is part of the total I(X,Y; S).

Williams and Beer introduced a new framework allowing the decomposition of mutual information between a system of predictors and a target variable (Williams and Beer, 2010). The framework proposes a decomposition of mutual information into non-negative terms that, including other beneficial properties, differentiate between redundant and synergistic contributions. This mathematical formalism is known as Partial Information Decomposition (PID). For simplicity, we will first present the theoretical framework for two predictors and a single target variable. Williams and Beer distinguished three types of contributions to the

information among two variables, or predictors, (here called A_1 and A_2 for distinction between the formalisms) and a third one, the target C (Suppl. Fig. 4). If A_1 and A_2 both carry the same information about C it suffices to know only one of them to fully determine the mutual information between them and C; this type of information is called *redundant* (Suppl. Fig. 4A). If only A_1 carries information about C it is referred to as *unique* information (Suppl. Fig. 4B); similarly, this holds true for A_2. Finally, if neither of the variables alone provide information about C, but they do only if observed together, we refer to it as *synergistic* information (Suppl. Fig. 4C). An exemplar case of synergistic information is the exclusive OR (XOR) operation (Suppl. Fig. 4C). In a XOR gate, both *unique* and *redundant* information are zero, whereas only *synergistic* effects carry information about C. In other words, the knowledge of any of the predictors separately does not provide any information about the target variable C. Only knowledge of both A_1 and A_2 together determines the value of C. In order to compute the different terms of the mutual information decomposition, either the synergy or the redundancy suffice, because the remaining terms can be thereafter computed. For example, if we compute the redundancy between A_1 and A_2, then the unique information carried by each source can be obtained by subtracting the redundancy from the mutual information between the particular source and the target variable $unique(A_1) = I(A_1; C) - redundancy(A_1, A_2)$. The mutual information $I(A_1; C)$ cannot contain any synergy, because the source A_1 is observed independently from A_2. Synergy thus becomes the remaining term of the full mutual information $I(A_1, A_2; C)$ after subtraction of all the other parts.

Within this formalism, Williams and Beer suggested a measure of redundancy called I_{min} which quantifies the information associated with a specific outcome c provided by a set of possible sources $A = \{A_1, A_2, ..., A_k\}$, and it is defined as:

$$I_{min}(C; \{A_1, A_2, ..., A_k\}) = \sum_c p(c)\, min_{A_i} I(C = c; A_i) \tag{11}$$

where $I(C = c; A_i)$ is a quantity called *specific information,* defined as:

$$I(C = c; A) = \sum_a p(a|c) \left[log\frac{1}{p(c)} - log\frac{1}{p(c|a)} \right] \tag{12}$$

Where a is one of all possible values of A_i. I_{min} quantifies redundancy in the form of information that is common to all the sources within A: the minimum information that any

source provides. The mutual information can be considered a special case of redundancy (i.e., the redundancy of a source with itself). This allows the use of the redundancy measures to compute all terms of the PID. However, the I_{min} measures themselves do not quantify the unique contribution of a single part of the decomposition. For example, $I_{min}(C;\{A_1\})$, which is equal to the mutual information between C and A_1, includes the redundant information shared by A_1 and A_2. Nevertheless, it is possible to reconstruct the information contributed independently by each source of the decomposition from the I_{min} measures, thanks to the natural ordering among all terms of the decomposition. For example, the redundancy between A_1 and A_2 would precede the unique contribution of either A_1 or A_2, because its knowledge is necessary in order to compute the unique contributions. This ordering creates a lattice from the elements of $A(R)$ (i.e., the set of all parts of the decomposition based on those sources). In the bivariate case, the terms are defined as follows: the redundant information between sources A_1 and A_2 is written as $\{A_1\}\{A_2\}$; the unique information of a source A_1 as $\{A_1\}$ and the synergy between A_1 and A_2 will be noted as $\{A_1, A_2\}$. For the decomposition of the mutual information between two sources A_1 and A_2 with a target C the lowest term in the lattice is the redundancy between the two sources, whereas the terms above include the unique contributions; the highest term incorporates the synergistic effects (Fig. 9A). Within this structure, a higher element in the lattice provides at least as much redundant information as any lower term; the highest element of the lattice then completes all the information that is present in the mutual information between all sources and the target $I(R;S)$. To compute the contribution of each node in the lattice, we need to subtract all information that is provided by underlying nodes. The amount of information provided by a single node is called partial information term $\prod_R$ and it is defined recursively as (see Williams and Beer 2010 for further details):

$$\prod_R(S;\alpha) = I_{min}(S;\alpha) - \sum_{\beta<\alpha} \prod_R(S;\beta) \tag{13}$$

We reasoned that we could exploit the PID formalism to put forward a feature-specific measure of information based on the decomposition of a system with a set of three predictors $R = \{X_{t-d}, Y_{t-d}, Y_t\}$ and its mutual information with the target stimulus feature S (Fig. 9B) . To satisfy the aforementioned third criterion, the novel measure should quantify the information about S present in Y_t that is redundant with respect to information already present in X_{t-d} and

unique with respect to Y_{t-d}. Consequently, the desired information measure could be expressed by the following partial information term (red dot in Fig. 9B):

$$FIT = \prod_{R} (S;\ \{X_{t-d}\}\{Y_t\}) \tag{14}$$

Note that it is not necessary to apply any conditioning on the past of Y, because this is implicitly accomplished by excluding the information captured by the lower $\prod_{R}$ term of the lattice that includes the past of Y, namely $\prod_{R}(S;\{X_{t-d}\}\{Y_t\}\{Y_{t-d}\}$.

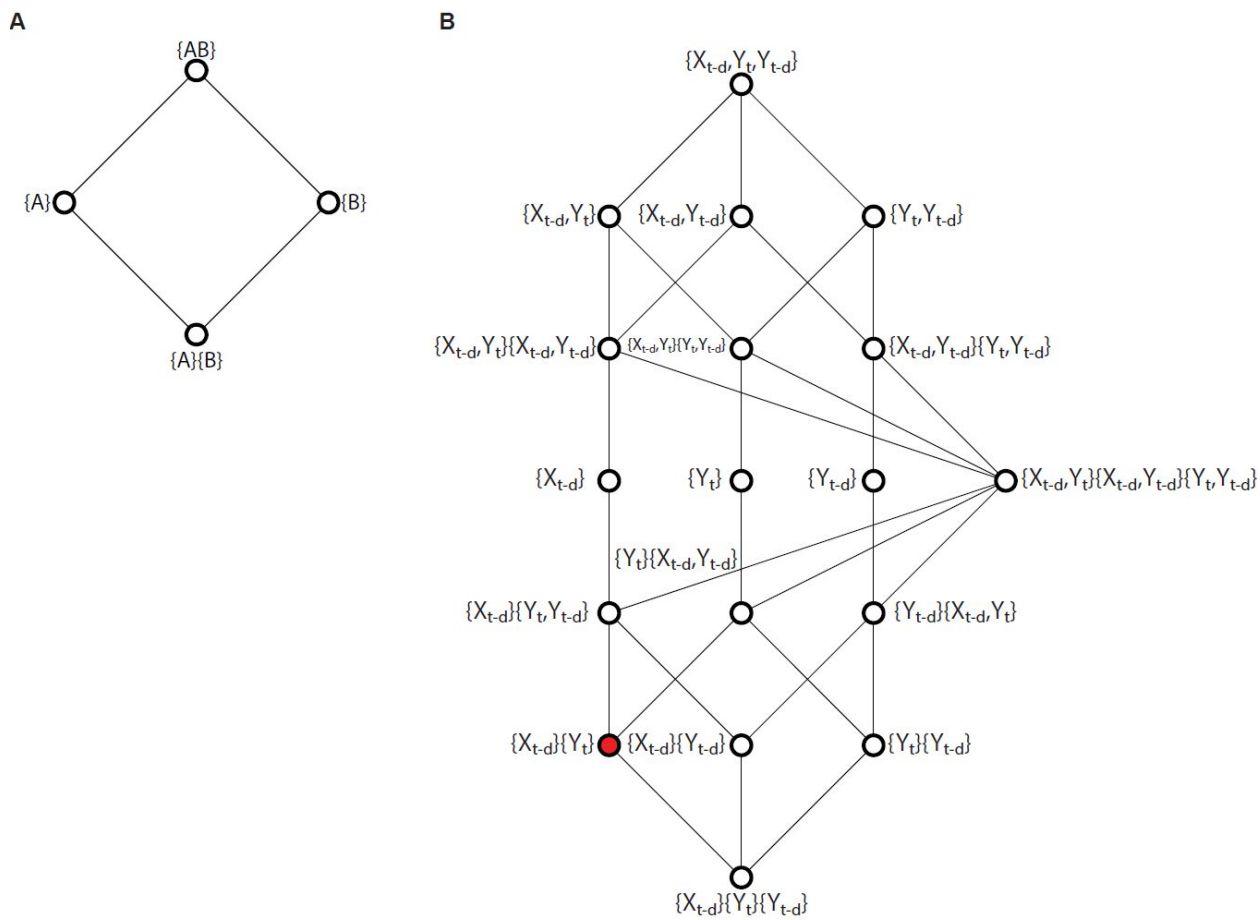

Figure 9. A Partial information decomposition lattice for 2 sources. **B** Partial information decomposition lattice for 3 source (X_{t-d}, Y_{t-d}, Y_t). The lattice is ordered based on the following: a collection of sources A is considered to "succeed" a second collection B (i.e., to be above in the lattice) if for each source in A there exists a source in B that does not provide any additional information than the one in A. The bottom term represents joint redundancy between all three variables and the top their synergy. The novel measure called FIT is marked in red.

Numerical simulations

We performed numerical simulations to test the reliability of our novel measure, the feature-specific information transfer (FIT). We analyzed different scenarios, each representing particular set of conditions under which feature-specific communication occurs.

Feature-specific communication was modeled by means of a stimulus-dependent transfer of information between two nodes, called X (i.e., the sender) and Y (i.e., the receiver of information). Each node carried a stimulus-dependent signal plus noise drawn from a Poisson distribution with mean equal to the signal value at a given time point. The presence of stimulus was simulated by adding a Gaussian profile with duration equal to the standard deviation of the Gaussian profile and centered at a time t_1. The amplitude of the Gaussian profile was then modulated by the stimulus value to an otherwise constant (i.e., baseline) signal X. The baseline activity of both X and Y was set to one. Information transfer from X to Y was constructed by adding the activity of X at time t_1 to signal Y at t_2. The time interval between t_1 and t_2 represented the communication delay. Stimulus values S, one per trial, were drawn from a uniform distribution and randomly chosen across trials. We used 4 possible values of stimulus.

Simulations were based on three scenarios. The first scenario (Fig. 1A, left panel) simulated a simple stimulus-dependent information transfer from node X to Y. The goal was to compare the behavior of FIT with respect to DI and DFI in quantifying stimulus-dependent communication. The aim of the second scenario (Fig. 1B, left panel) was to study information transfer without a stimulus-dependent modulation. The aim was to study the ability of FIT to efficiently dissociate between stimulus-dependent and stimulus-independent transfer. To do so, the sender node X was split into two sub-nodes, called X^+ and X^-. X^+ was simulated as in scenario 1, whereas X^- was equal to baseline activity minus the Gaussian profile of signal X^+ plus a noise drawn from a Poisson distribution with mean equal to the signal value at a given time point. The baseline activity was set to the constant part of the signal, where the stimulus representing Gaussian profile was not added and it was identical for both X^+ and X^-. Activities of both sub-nodes were then added together and normalized before being injected into the signal of the receiver Y. This process effectively destroys stimulus-dependent information transfer to signal Y, while leaving stimulus-independent transfer unaffected. The third scenario (Fig. 1C, left panel) was designed to study the effect of an external source of information influencing signal Y without passing through the communication channel between X to . To

do so, we added information about a second stimulus S' having the same probability distribution as , but drawn from it independently, to signal Y. The Gaussian profile added to the signal Y at t_2 was thus modulated by the stimulus values S'. All simulations were run for 120 timestamps and mimicked the evolution of activity of two hypothetical neural populations over a time interval from -20 to 100 msec across several repetitions (i.e., experimental trials). For each scenario, we simulated 100 000 trials, except for those used for robustness testing. Robustness tests were run on 2^8 to 2^{17} number of trials.

Information theoretical analysis

For every simulation, we computed values of DI, DFI and FIT in a time-resolved manner. To compute the full joint probability distribution required for the calculation of information theoretical measures, we binned all variables at a given time point across trials into three bins, using a technique that distributes equally the number of samples per bin. The number of bins was set to three, as a trade-off between the size of the joint probability distribution and computability cost. For the simulations showing robustness to noise, effects of conditioning and high synergy, we took all past values of X and Y from a time interval $-d$, where d stands for a delay. The optimal delay was set to the distance between peaks of the stimulus-related activity. We additionally computed the information quantities at each time point for a set of delays in interval ranging from 1 to 45 samples and then took the average of those values. However, in the scenario showing the effects of conditioning with respect to different delays, we computed information measures for $d = 10$, 30 and 50 samples (Fig. 3). In order to correct for biases due to the limited number of samples, we used the Quadratic Extrapolation procedure (Panzeri et al., 2007). The bias grows quadratically as a function of the logarithmic decrease in the number of trials. To correct for such bias, we first computed the value of a given information quantity using either all, half or a quarter of the available simulated trials, and then we fitted a quadratic curve. The constant coefficient of the quadratic curve was the corrected value for the information quantity. To increase accuracy in the information theoretic measures, we computed the values for both halves and all quarters and then averaged across them.

Statistical analysis

In order to assess the level of significance of the different information theoretic measures, we used nonparametric permutation techniques. For every time point, we established its 0.1th – 99.9th quantile interval under the null hypothesis. To obtain the values under the null

hypothesis, we randomly shuffled stimulus order across trials for DFI and FIT, and values of X for DI. In this manner, we computed 100 surrogate values based on independent permutations. To establish the desired percentile values, we then fitted a Gaussian distribution through the surrogates.

Analysis of Neurophysiological Data

In order to demonstrate the relevance of our measure in brain communication analysis, we analyzed three neurophysiological datasets. The goal of the first analysis was to compare the different metrics (DI, DFI and FIT) in their ability to quantify stimulus-dependent information flow between EEG signals. To do so, we analyzed an EEG dataset in which participants have to perform a face detection task (Rousselet et al., 2014). The goal of the second analysis was to compare the different metrics (DI, DFI and FIT) in their ability to quantify stimulus-dependent information flow between MUA signals recorded from rat's brain in a multisensory stimulations scenario (visual and/or tactile stimulation) (Bieler et al., 2018). The goal of the third analysis was to assess the dynamics of visuomotor-related (i.e., event-related *versus* baseline) modulations in DI, DFI and FIT as they evolve over time. To do so, we exploited an MEG dataset containing source-level high-gamma activity from human participants performing a visuomotor mapping task (Brovelli et al., 2017, 2015).

EEG and face detection task

Experimental conditions and behavioral tasks

In the first set of analyses, we computed the DI, DFI and FIT measures on a publicly available EEG dataset (Rousselet et al., 2014). The data was recorded during a face detection task. Participants (N=16) were presented with an image hidden behind a bubble mask. On half of the trials, the image was a face and the other half contained a random texture. Participants were instructed to determine the content of the image and report whether a face was present or not. For our purposes, we only considered correct trials where the face was correctly detected by the participants (approximately 1000 trials per subject). Previous studies have shown that presence of an eye is one of the most important aspects for determining whether an image contains a face or not. Indeed, it modulates reaction times and the latency of the N170 response in lateral temporo-occipital EEG electrodes (Rousselet et al., 2014). More recently, Ince and colleagues (Ince et al., 2015) showed that temporo-occipital electrodes contain information about the contralateral eye later in time. The aim of our analyses was to decode the presence of

an eye from both occipito-temporal sensors in order to determine the origin of that information. The dataset contains preprocessed EEG data from the electrodes in the left and the right occipito-temporal regions that display the highest mutual information with the visibility of the contralateral eye. The dataset contains 15 subjects as one was discarded for low quality of the recording.

Information theoretic measures

We computed the first derivatives of the EEG signal for both occipito-temporal sensors and used both its absolute values and first derivatives to compute the information quantities. We used both the raw and absolute time derivatives to be consistent with the analyses performed in previous papers. To establish the joint probability distribution for computation of the information quantities, we binned both the derivatives and absolute values into 3 equally populated bins resulting into 9 possible values for each random variable. The visibility of an eye was considered as the stimulus and it was also binned into 3 bins. We computed the information quantities for all combinations of direction of the transfer (left to right, right to left) and a particular eye (left and right). These values were computed for every subject and then we averaged across them.

Statistical analyses

We established significant increase in information measures using a cluster-based nonparametric statistical test introduced in (Maris and Oostenveld, 2007). First, we computed the surrogate values of the information quantities for all time points and delays using 100 permutations of eye visibility. For each of those permutations, we created clusters of values higher than a threshold, based on 8-adjacency in the 2D time-delay grid. We set the threshold to 97.5^{th} quantile, that we obtained by ranking all the values in the grid. Then, we computed cluster level statistic by summing all values of the given information quantity within the cluster. Finally, we took the largest cluster-level statistic for each of the permutations and compared them to cluster-level statistic in the non-shuffled data. We considered as significant those clusters that had a cluster-level statistic smaller than 5 clusters obtained from the surrogates.

Robustness with respect to the sample size bias

We compared values of sample size bias corrected and uncorrected information quantities under two conditions, the information transfer being present in the data or not. To compute values for each condition, we determined in which areas of two-dimensional grid of all

combination of time and delay for which we computed the information quantities the transfer was and was not present. The representative value of the first condition was an average across all values in an area with the presence of transfer and the representative value of the second condition was an average across values in an identical area moved in time in such manner that it only encapsulated points with no transfer. We computed these values for both conditions in simulations that had numbers of trials from 2^8 to 2^{18} and each simulation was repeated 5 times across which we averaged all the computed values.

MUA and multisensory stimulation

Experimental conditions

In the second set of analyses, we compared the three different metrics (DI, DFI and FIT) on multiunit activity (MUA) recorded from rats (N=6). The data was recorded from the visual and somatosensory territories of the thalamus and corresponding primary cortical regions in three stimulation conditions: visual stimulation, whiskers tactile stimulation and bimodal stimulation (simultaneous visual and tactile) (Bieler et al., 2018). The aim of our analyses was to assess the presence of stimulus-specific information transfer from thalamus to cortex. In particular, we investigated the ability of our novel measure to characterize the functional properties of two neural circuits: 1) the visual pathway carrying the visual information from the eyes to the primary visual cortex (V1) through the Lateral Geniculate Nucleus (LGN) of the thalamus; 2) the somatosensory pathway carrying the sensory information from the whiskers receptors to the primary somatosensory cortex (S1) through the Ventral Posteromedial nucleus (VPM) of the thalamus. All experiments were conducted during the light phase under sleep-like conditions mimicked by urethane anesthesia (Bitzenhofer et al., 2015). By these means, the interference with spontaneous whisking and the impact of alert state, which modulate cross-modal integration, were avoided. Unimodal (either light flash or whisker deflection) or bimodal (simultaneous light flash and whisker deflection) stimuli were applied using a custom-made stimulation device as described in (Sieben et al., 2015, 2013). Briefly, whiskers were stimulated by deflection through compressed air-controlled roundline cylinders gated via solenoid valves. The device produced almost silent (6~10dB), non-electrical stimulation with precise timing (0.013 ± 0.81 ms) that was constant over all trials/conditions. For full eye field visual stimulation, 50-ms-long LED light flashes (300 lux) were used. For bimodal stimulation, whisker deflection and light flashes were applied in the same hemifield. Stimuli were randomly presented across trials in blocks of three different stimulation conditions (unimodal tactile, unimodal visual,

bimodal visual-tactile). In our analysis we considered only stimulation contralateral to the recording brain areas. Each type of stimulus was presented 100 times contralateral to the recording electrodes with an interstimulus interval of 6.5 ± 0.5 s. To achieve a physically simultaneous stimulation of whiskers (valve-controlled whisker stimulation) and eyes (instantaneous light flash), the time delay of whisker stimulation was calculated to match visual stimulation onset. The non-stimulated eye was covered with an aluminum foil patch.

The surgery was performed under ketamine/Xylazine anesthesia, the rat's eyes were covered with ointment (Bepanthen), and the ear canals were filled with silicon adhesive (Kwik-Sil, World Precision Instruments) to block auditory input. Extracellular recordings of the local field potential (LFP) and multiunit activity (MUA) were performed from head-fixed rats under light urethane anesthesia (0.5 g/kg body weight, i.p.; Sigma-Aldrich) using custom made one-shank 1632-channel electrodes (0.5-3 M_Ω Silicon Michigan probes, Neuronexus Technologies; 100-_µm intersite spacing) that were angularly inserted into S1 barrel field (ventral-medial, 50°, 2.4-2.6 mm posterior and 5.5-5.8 mm lateral to bregma) and V1 (ventral-rostral, 45°, 6.9-7.1 mm posterior and 3.4-3.7 mm lateral to bregma) to a depth of 5.3 mm and 5.5 mm, respectively. The placement of recording sites enabled simultaneous recording from supragranular (S), granular (G) and infragranular (I) layers of S1 and the ventral posteromedial nucleus (VPM) of the thalamus as well from S, G and I layers of V1 and the dorsal geniculate nucleus (dLGN) of the thalamus. Electrodes spanned supragranular, granular, and infragranular layers (please add VPM and LGN) and were labeled with DiI (1,1'-dioctadecyl-3,3,3',3'- tetramethylindocarbocyanine; Invitrogen) for postmortem reconstruction of their tracks in histologic sections. A silver wire was inserted into the cerebellum and served as ground and reference electrode. The body temperature of the animal was kept constant at 37°C during recording. The position of recording sites over layers was confirmed by electrophysiological (i.e., reversal of the evoked potential between supragranular and granular layers) and histologic (i.e., granular cell body layer) landmarks. Neural activity was recorded at a sampling rate of 32 kHz using a multichannel extracellular amplifier (no gain, Digital Lynx 10S, Neuralynx) and the acquisition software Cheetah. The signal was bandpass filtered (0.1 Hz and 5 kHz) by the Neuralynx recording system, for antialiasing, and then down-sampled by a factor of 8 obtaining a sampling rate of 8 kHz (Bieler et al., 2018). In the current work, we used the recordings from infragranular layers of S1 and V1 and from VPM and LGN. Data was imported and analyzed offline using custom-written tools in Matlab software (version R2018A (Math-Works), and Multi Unit Activity (MUA) was extracted (Safaai et al., 2015). For all

channels, spike times were first detected from the band-passed (400–3,000 Hz, fourth-order IIR Butterworth Filter) extracellular potential in each electrode by threshold crossing (>3 SD). A spikes train was obtained for all channels using a temporal binning of 0.125 ms (1/8kHz). For each brain region, mass MUA was obtained by pooling together the trains of all recorded spikes form all electrodes related to that region, a refractory period of 1ms was considered and the resulting train was resampled at 1 KHz (1 ms time bins).

Information theoretic measures and statistical analysis

Information theoretical analyses were performed on the MUA signals recorded for LGM, VPM, V1 and S1. To establish the joint probability distribution for computation of the information quantities, we computed the spiking rate for 10 msec temporal bins and then we binned the values into 2 equally populated bins resulting into 2 possible values for each random variable. We considered two possible stimulation conditions:

1. Condition 1 = {V, V+T} – unimodal visual stimulation and bimodal stimulation (simultaneous visual and tactile)
2. Condition 2 = {T, V+T} – unimodal tactile stimulation and bimodal stimulation (simultaneous visual and tactile)

In condition 1, the difference between the two types of stimulation is due to the presence of the tactile stimulation; our analyses aimed at testing the hypothesis that tactile-specific information transfer flows along the neural pathway selective for tactile features (i.e., the VPM to S1 pathway). On the contrary, in condition 2 the difference between stimulations types is due to the visual stimulation; therefore, we expected to observe visual-specific information transfer occurs along the neural pathway selective for visual stimulation (i.e. the LGN to V1 pathway). We computed the information quantities (DI, DFI, FIT) for all combinations of directions of the transfer (from V1 to LGN and vice-versa; from S1 to VPM and vice-versa) and a particular stimulus set. These values were computed for every subject and then we averaged across them. The statistical analysis was done as described in the previous EEG section.

Source-level high-gamma MEG activity and visuomotor task

Experimental conditions, behavioral tasks and brain data acquisition

The third set of analyses was performed on an MEG dataset collected while participants performed an associative visuomotor mapping task (Brovelli et al., 2017, 2015). The task required participants to perform a finger movement associated to a digit number: digit “1”

instructed the execution of the thumb, “2” for the index finger, “3” for the middle finger and so on. Maximal reaction time was 1s. After a fixed delay of 1 second following the disappearance of the digit number, an outcome image was presented for 1 s and informed the subject whether the response was correct, incorrect, or too late (if the reaction time exceeded 1 s). Incorrect and late trials were excluded from the analysis, because they were either absent or very rare (i.e., maximum 2 late trials per session). The next trial started after a variable delay ranging from 2 to 3 s (randomly drawn from a uniform distribution) with the presentation of another visual stimulus. Each participant performed two sessions of 60 trials each (total of 120 trials). Each session included three digits randomly presented in blocks of three trials. The average reaction time was 0.504s ± 0.004s (mean ± s.e.m.). Anatomical T1-weighted MRI images were acquired for all participants using a 3-T whole-body imager equipped with a circular polarized head coil. Magnetoencephalographic (MEG) recordings were performed using a 248 magnetometers system (4D Neuroimaging magnes 3600). Visual stimuli were projected using a video projection and motor responses were acquired using a LUMItouch® optical response keypad with five keys. Presentation® software was used for stimulus delivery and experimental control during MEG acquisition.

Brain Atlas: MarsAtlas

Single-subject cortical parcellation was performed using the *MarsAtlas* brain scheme (Auzias et al., 2016). After denoising using a non-local means approach (Coupé et al., 2008), T1-weighted MR-images were segmented using the FreeSurfer “recon-all” pipeline (http://freesurfer.net). Grey and white matter segmentations of each hemisphere were imported into the BrainVisa software and processed using the Morphologist pipeline procedure (http://brainvisa.info). White matter and pial surfaces were reconstructed and triangulated, and all sulci were detected and labeled automatically (Mangin et al., 2004; Perrot et al., 2011). A parameterization of each hemisphere white matter mesh was performed using the Cortical Surface toolbox (http://www.meca-brain.org/softwares/). It resulted in a 2D orthogonal system defined on the white matter mesh, constrained by a set of primary and secondary sulci (Auzias et al., 2013). The parcels corresponding to the subcortical structures were extracted using Freesurfer (Fischl et al., 2002). The subcortical structures included in the brain parcellation were the caudate nucleus, putamen, nucleus accumbens, globus pallidus, thalamus, amygdala, and hippocampus. The whole-brain parcellation therefore comprised 96 areas (41 cortical and 7 subcortical areas per hemisphere).

Single-trial high-gamma activity (HGA) in MarsAtlas

MEG signals were down-sampled to 1 kHz, low-pass filtered to 250 Hz and segmented into epochs aligned on finger movement (i.e., button press). Epoch segmentation was also performed on stimulus onset and the data from -0.5 and -0.1 s prior to stimulus presentation was taken as baseline activity for the calculation of the single-trial high-gamma activity (HGA). Artefact rejection was performed semi-automatically and by visual inspection. For each movement-aligned epoch and channel, the signal variance and z-value were computed over time and taken as relevant metrics for the identification of artefact epochs. All trials with a variance greater than 1.5*10-24 across channels were excluded from further search of artefacts. Metrics such as the z-score, absolute z-score, and range between the minimum and maximum values were also inspected to detect artefacts. Channels and trials displaying outliers were removed. Two MEG sensors were excluded from the analysis for all subjects.

Spectral density estimation was performed using multi-taper method based on discrete prolate spheroidal (slepian) sequences (Mitra and Pesaran, 1999; Percival and Walden, 1993). To extract high-gamma activity from 60 to 120Hz, MEG time series were multiplied by *k* orthogonal tapers (*k* = 8) (0.15s in duration and 60Hz of frequency resolution, each stepped every 0.005s), centered at 90Hz and Fourier-transformed. Complex-valued estimates of spectral measures $X^{n}_{sensor}(t,k)$, including cross-spectral density matrices, were computed at the sensor level for each trial *n*, time *t* and taper *k*. Source analysis requires a physical forward model or leadfield, which describes the electromagnetic relation between sources and MEG sensors. The leadfield combines the geometrical relation of sources (dipoles) and sensors with a model of the conductive medium (i.e., the head model). For each participant, we generated a headmodel using a single-shell model constructed from the segmentation of the cortical tissue obtained from individual MRI scans as described in section 3.2 (Nolte, 2003). Leadfields were not normalized. Sources were placed in the single-subject volumetric parcellation regions. For each region, we computed the number of sources *nSP* as the ratio of the volume and the volume of a sphere of radius equal to 3 mm. The k-means algorithm (Tou and González, 1974) was then used to partition the 3D coordinates of the voxels within a given volumetric region into *nS* clusters. The sources were placed at the center of each partition for each brain region. The head model, source locations and the information about MEG sensor position for both models were combined to derive single-participant leadfields. The orientation of cortical sources was set perpendicular to the cortical surface, whereas the orientation for subcortical sources was left

unconstrained.

We used adaptive linear spatial filtering (Van Veen et al., 1997) to estimate the power at the source level. In particular, we employed the Dynamical Imaging of Coherent Sources (DICS) method, a beamforming algorithm for the tomographic mapping in the frequency domain (Groß et al., 2001), which is a well suited for the study of neural oscillatory responses based on single-trial source estimates of band-limited MEG signals (for a series of review see, (Hansen et al., 2010)). At each source location, DICS employs a spatial filter that passes activity from this location with unit gain while maximally suppressing any other activity. The spatial filters were computed on all trials for each time point and session, and then applied to single-trial MEG data. DICS allows the estimate of complex-value spectral measures at the source level, $X^n_{source}(t,k) = A(t)X^n_{sensor}(t,k)$, where $A(t)$ is the spatial filter that transforms the data from the sensor to source level and $X^n_{sensor}(t,k)$ is the complex-valued estimates of spectral measures, including cross-spectral density matrices, computed at the sensor level for each trial *n*, time *t* and taper *k* (for a detailed description of a similar approach see (Hipp et al., 2011)). The single-trial high-gamma power at each source location was estimated by multiplying the complex spectral estimates with their complex conjugate, and averaged over tapers *k*, $P^n_{source}(t) = \langle X^n_{source}(t,k)X^n_{source}(t,k)^* \rangle_k$, where angle brackets refer to the average across tapers and * to the complex conjugate. Single-trial power estimates aligned on movement and stimulus onset were log-transformed to make the data approximate Gaussian and low-pass filtered at 50Hz to reduce noise. Single-trial mean power and standard deviation in a time window from -0.5 and -0.1 s prior to stimulus onset was computed for each source and trial, and used to z-transform single-trial movement-locked power time courses. Similarly, single-trial movement-locked power time courses were log-transformed and z-scored with respect to baseline period, so to produce HGAs for the prestimulus period from -1.6 to -0.1 s with respect to stimulation for subsequent functional connectivity analysis. Finally, single-trial HGA for each brain region of *MarsAtlas* was computed as the mean z-transformed power values averaged across all sources within the same region.

Information theoretic measures

The goal of the MEG analysis was to assess the dynamics of visuomotor-related modulations in DI, DFI and FIT as they evolve over time. The DFI and FIT were thus computed by taking as feature the visuomotor-related *versus* baseline HGA. The visuomotor-related activity was considered as the HGA aligned to movement onset (from -1s to 0.75s), whereas the baseline activity was defined as the HGA at -0.5s prior to stimulus onset. To estimate the joint probability distribution at each time stamp around movement onset, we concatenated the HGA in windows of 20 msec (4 time instants), trials and subjects, and we binned the values into three equally-populated bins, thus resulting into 6 possible values for each random variable. The DI, DF and FIT measures were thus computed in a time-resolved manner on all time points aligned on motor response. We computed information theoretic measures to quantify the information transfer about visuomotor processing between three regions of interest: the medial superior parietal cortex (SPCm), the dorsolateral premotor cortex (PMdl) and the dorsal Inferior Parietal Cortex (IPCd).

References

Amblard P-O, Michel OJJ. 2011. On directed information theory and Granger causality graphs. *Journal of Computational Neuroscience*. doi:10.1007/s10827-010-0231-x

Auzias G, Coulon O, Brovelli A. 2016. MarsAtlas: A cortical parcellation atlas for functional mapping. *Hum Brain Mapp* **37**:1573–1592.

Auzias G, Lefevre J, Le Troter A, Fischer C, Perrot M, Régis J, Coulon O. 2013. Model-driven harmonic parameterization of the cortical surface: HIP-HOP. *IEEE Trans Med Imaging* **32**:873–887.

Bell AJ. 2003. The co-information lattice. Presented at the Proceedings of the Fifth International Workshop on Independent Component Analysis and Blind Signal Separation: ICA.

Bertschinger N, Rauh J, Olbrich E, Jost J, Ay N. 2014. Quantifying unique information. *Entropy* **16**:2161–2183.

Bieler M, Xu X, Marquardt A, Hanganu-Opatz IL. 2018. Multisensory integration in rodent tactile but not visual thalamus. *Sci Rep* **8**:15684.

Bitzenhofer SH, Sieben K, Siebert KD, Spehr M, Hanganu-Opatz IL. 2015. Oscillatory activity in developing prefrontal networks results from theta-gamma-modulated synaptic inputs. *Cell Rep* **11**:486–497.

Bressler SL, Menon V. 2010. Large-scale brain networks in cognition: emerging methods and principles. *Trends Cogn Sci* **14**:277–290.

Bressler SL, Seth AK. 2011. Wiener–Granger causality: a well established methodology. *Neuroimage* **58**:323–329.

Brovelli A, Badier J-M, Bonini F, Bartolomei F, Coulon O, Auzias G. 2017. Dynamic reconfiguration of visuomotor-related functional connectivity networks. *Journal of Neuroscience* **37**:839–853.

Brovelli A, Chicharro D, Badier J-M, Wang H, Jirsa V. 2015. Characterization of cortical networks and corticocortical functional connectivity mediating arbitrary visuomotor mapping. *Journal of Neuroscience* **35**:12643–12658.

Brovelli A, Ding M, Ledberg A, Chen Y, Nakamura R, Bressler SL. 2004. Beta oscillations in a large-scale sensorimotor cortical network: directional influences revealed by Granger causality. *Proc Natl Acad Sci U S A* **101**:9849–9854.

Corbetta M, Shulman GL. 2002. Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci* **3**:201–215.

Coupé P, Yger P, Prima S, Hellier P, Kervrann C, Barillot C. 2008. An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images. *IEEE Trans Med Imaging* **27**:425–441.

Culham JC, Valyear KF. 2006. Human parietal cortex in action. *Curr Opin Neurobiol* **16**:205–212.

Ding M, Chen Y, Bressler SL. 2006. 17 Granger causality: basic theory and application to neuroscience. *Handbook of time series analysis: recent theoretical developments and applications* **437**.

Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, van der Kouwe A, Killiany R, Kennedy D, Klaveness S, Montillo A, Makris N, Rosen B, Dale AM. 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* **33**:341–355.

Granger CWJ. 1980. Testing for causality: a personal viewpoint. *J Econ Dyn Control* **2**:329–352.
Griffith V, Koch C. 2014. Quantifying Synergistic Mutual Information In: Prokopenko M, editor. Guided Self-Organization: Inception. Berlin, Heidelberg: Springer Berlin Heidelberg. pp. 159–190.
Groß J, Kujala J, Hämäläinen M, Timmermann L, Schnitzler A, Salmelin R. 2001. Dynamic imaging of coherent sources: studying neural interactions in the human brain. *Proceedings of the National Academy of Sciences* **98**:694–699.
Hansen P, Kringelbach M, Salmelin R. 2010. MEG: An introduction to methods. Oxford university press.
Harder M, Salge C, Polani D. 2013. Bivariate measure of redundant information. *Physical Review E* **87**:012130.
Hipp JF, Engel AK, Siegel M. 2011. Oscillatory synchronization in large-scale cortical networks predicts perception. *Neuron* **69**:387–396.
Ince RAA, Jaworska K, Gross J, Panzeri S, Van Rijsbergen NJ, Rousselet GA, Schyns PG. 2016. The Deceptively Simple N170 Reflects Network Information Processing Mechanisms Involving Visual Feature Coding and Transfer Across Hemispheres. *Cereb Cortex* **26**:4123–4135.
Ince RAA, Van Rijsbergen NJ, Thut G, Rousselet GA, Gross J, Panzeri S, Schyns PG. 2015. Tracing the flow of perceptual features in an algorithmic brain network. *Sci Rep* **5**:17681.
Mangin J-F, Riviere D, Cachia A, Duchesnay E, Cointepas Y, Papadopoulos-Orfanos D, Scifo P, Ochiai T, Brunelle F, Regis J. 2004. A framework to study the cortical folding patterns. *Neuroimage* **23**:S129–S138.
Maris E, Oostenveld R. 2007. Nonparametric statistical testing of EEG-and MEG-data. *J Neurosci Methods* **164**:177–190.
Massey J. 1990. Causality, feedback and directed information. Presented at the Proc. Int. Symp. Inf. Theory Applic.(ISITA-90), Citeseer.
McGill W. 1954. Multivariate information transmission. *Transactions of the IRE Professional Group on Information Theory* **4**:93–111.
Mitra PP, Pesaran B. 1999. Analysis of dynamic brain imaging data. *Biophys J* **76**:691–708.
Nolte G. 2003. The magnetic lead field theorem in the quasi-static approximation and its use for magnetoencephalography forward calculation in realistic volume conductors. *Phys Med Biol* **48**:3637.
Panzeri S, E R, Montemurro MA, Petersen RS. 2007. Correcting for the sampling bias problem in spike train information measures. *J Neurophysiol* **98**:1064–1072.
Panzeri S, Treves A. 1996. Analytical estimates of limited sampling biases in different information measures. *Network: Computation in Neural Systems* **7**:87–107.
Percival DB, Walden AT. 1993. Spectral analysis for physical applications. Cambridge University Press.
Perrot M, Rivière D, Mangin J-F. 2011. Cortical sulci recognition and spatial normalization. *Med Image Anal* **15**:529–550.
Pola G, Thiele A, Hoffmann KP, Panzeri S. 2003. An exact method to quantify the information transmitted by different mechanisms of correlational coding. *Network-Computation in Neural Systems* **14**:35–60.
Rousselet GA, Ince RAA, van Rijsbergen NJ, Schyns PG. 2014. Eye coding mechanisms in early human face event-related potentials. *J Vis* **14**:7–7.
Safaai H, Neves R, Eschenko O, Logothetis NK, Panzeri S. 2015. Modeling the effect of locus coeruleus firing on cortical state dynamics and single-trial sensory processing. *Proc Natl*

Acad Sci U S A **112**:12834–12839.
Schreiber T. 2000. Measuring Information Transfer. *Phys Rev Lett* **85**:461–464.
Seth AK, Barrett AB, Barnett L. 2015. Granger causality analysis in neuroscience and neuroimaging. *Journal of Neuroscience* **35**:3293–3297.
Shannon CE. 1948. A mathematical theory of communication. *The Bell System Technical Journal* **27**:379–423.
Sieben K, Bieler M, Röder B, Hanganu-Opatz IL. 2015. Neonatal Restriction of Tactile Inputs Leads to Long-Lasting Impairments of Cross-Modal Processing. *PLoS Biol* **13**:e1002304.
Sieben K, Röder B, Hanganu-Opatz IL. 2013. Oscillatory entrainment of primary somatosensory cortex encodes visual control of tactile processing. *J Neurosci* **33**:5736–5749.
Strong SP, Koberle R, van Steveninck RR de R, Bialek W. 1998. Entropy and information in neural spike trains. *Phys Rev Lett* **80**:197.
Tou JT, González RC. 1974. Pattern recognition principles. Addison-Wesley Pub. Co.
Van Veen BD, Van Drongelen W, Yuchtman M, Suzuki A. 1997. Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Transactions on biomedical engineering* **44**:867–880.
Varela F, Lachaux J-P, Rodriguez E, Martinerie J. 2001. The brainweb: phase synchronization and large-scale integration. *Nat Rev Neurosci* **2**:229–239.
Vicente R, Wibral M, Lindner M, Pipa G. 2011. Transfer entropy—a model-free measure of effective connectivity for the neurosciences. *J Comput Neurosci* **30**:45–67.
von der Malsburg C, Phillips WA, Singer W. 2010. Dynamic coordination in the brain: from neurons to mind. MIT Press.
Wibral M, Vicente R, Lizier JT. 2014. Directed information measures in neuroscience. Springer.
Williams PL, Beer RD. 2010. Nonnegative decomposition of multivariate information. *arXiv preprint arXiv:10042515*.
Wise SP, Murray EA. 2000. Arbitrary associations between antecedents and actions. *Trends Neurosci* **23**:271–276.
Wise SP, Murray EA, Gerfen CR. 1996. The frontal cortex-basal ganglia system in primates. *Crit Rev Neurobiol* **10**:317–356.