Neuron-level dynamics of oscillatory network structure and markerless tracking of kinematics during grasping

Dissertation

For the awarding of the degree

"Doctor rerum naturalium"

of Georg-August-Universität Göttingen

within the doctoral program Theoretical and Computational Neuroscience

of the Georg-August University School of Science (GAUSS)

submitted by

Swathi Sheshadri

from Hassan, Karnataka, India

Göttingen, 2020

Doctoral thesis committee

Prof. Dr. med Hansjörg Scherberger (First Referee, Supervisor)
Neurobiology Laboratory
German Primate Center
Kellnerweg 4
37077 Göttingen

Prof. Dr. Alexander Gail (Second Referee)
Cognitive Neuroscience Laboratory
German Primate Center
Kellnerweg 4
37077 Göttingen

Prof. Dr. Michael Wibral
Campus Institute for Dynamics of Biological Networks
Department of Data-driven Analysis of Biological Networks
Kellnerweg 6
37077 Göttingen

Additional members of the examination board

Prof. Dr. Melanie Wilke
Department of Cognitive Neurology
University Medical Center Göttingen
Robert-Koch-Straße 40
37075 Göttingen

Prof. Dr. Tim Gollisch
Sensory Processing in the Retina
University Medical Center Göttingen
Waldweg 33
37073 Göttingen

Dr. Igor Kagan,
Decision and Awareness Group,
Cognitive Neuroscience Laboratory,
German Primate Center,
Kellnerweg 4,
37077 Göttingen

Date of oral examination: December 1th, 2020

Herewith I declare that I have written this thesis independently and with no other aids and sources other than quoted.

Göttingen, 30th of September, 2020 Swathi Sheshadri

To my parents

Acknowledgments

I am very thankful to Hans Scherberger, for giving me the opportunity to pursue research in the fascinating field of Neuroscience. Throughout my study, he has provided me ideas, encouragement, support and freedom to choose and work on projects of my interest. Hans's motivating and composed nature has ensured a great environment in the lab, allowing students to greatly enjoy the research experience. I would also like to thank the other thesis committee members Alexander Gail and Michael Wibral, for their helpful feedback and constructive criticism during the course of my study.

Thank you to all the members of the Neurobiology lab. Thanks especially to Benjamin Dann for mentoring me and helping me navigate well through the PhD landscape. Ben has been a rich source of advice, be it related to scientific or any other matter. The office that I shared with Timo Hueser, Andrej Filippow and Ben gave me a chance to participate in many cherishable friendly banters and scientific discussions, for which I am thankful.

I also want to thank Matthias Dörge, Natalie Bobb, Sabine Borchert, Ricarda Lbik, Anja Osterbrink and Kerstin Zaft for their support in technical, administrative and animal training related work. Special thanks to Matthias Dörge for his patience in helping me improve my German language skills.

I am also thankful to some of my previous teachers and mentors, particularly Suma, Anuradha, Aprameya Swarup, Rangaswamy and Ignacio who have guided me during my early academic years.

I thank my family. My husband Shreyas, for always looking out for me and for being so full of pleasant surprises. Thanks also for his criticism, which has been constructive and helped me improve. My little sister Smitha, for being the wisest person in our family. My parents, for always being there for me, always believing in me, and for instilling a love for learning in me.

Table of Contents

1 General Introduction 11

1.1 Neuronal signal recording techniques **13**

1.2 Functional connectivity **18**

1.2.1 Field-field synchrony 20

1.2.2 Spike-spike and spike-field synchrony 22

1.3 Selective and flexible information processing **25**

1.4 The fronto-parietal grasping network **31**

1.5 Tracking hand kinematics **36**

1.5.1 Instrumented data gloves 38

1.5.2 Camera based hand tracking systems 40

1.5.3 Deep learning for computer vision 42

1.5.4 ConvNets in hand tracking applications 45

2 Original Articles and Manuscripts 49

2.1 Distinct beta and low frequency oscillatory neurons participate in a behavior dependent reconfiguration of the fronto-parietal grasping network in macaques. **50**

2.2 Precise markerless tracking of the hand during object grasping **103**

2.3 3D reconstruction toolbox for behavior tracked with multiple cameras **132**

3 General Discussion 144

3.1 Summary **144**

3.2 Outlook **146**

4 Bibliography 150

5 Appendix A : Other co-authored manuscript **167**

6 Appendix B : Supplementary tables **195**

7 Curriculum Vitae 199

1 General Introduction

We select and listen to our favorite channels by quickly tuning the radio into specific frequencies. Oscillators, which are an important part of the electronic circuitry of the radio, are not only limited to man-made gadgets but also occur frequently in biological systems. For instance, the central pattern generators in the spinal cord facilitate rhythmic motor activity during behaviors such as walking, feeding and swimming (Kandel *et al.*, 2000). In the human brain, oscillatory activity was first observed by Hans Berger in the 1920s using the Electroencephalography (EEG) technique that he developed (Millett, 2001; Buzsáki, 2005). EEG non-invasively records field potential signals that mainly reflect summed synaptic activity (transmembrane currents) in local neuronal populations (Buzsáki, Anastassiou and Koch, 2012; Pesaran *et al.*, 2018; Schneider *et al.*, 2020). In the seminal paper on EEG (Berger, 1929), Hans Berger observed two types of waves in the EEG recordings. First, higher amplitude waves that appeared 10-11 times in one second, which he referred to as alpha waves. Second, smaller waves that appeared 20-30 times in one second, which he referred to as beta waves. Furthermore, the alpha waves during his experiments were observed when the subjects closed their eyes and the beta waves were observed when the subjects kept their eyes open, suggesting an association between oscillations and behavior. In agreement with this, numerous studies have demonstrated the correlation between oscillations at different frequencies with different behaviors, referred to as the behavioral correlates of oscillations. To investigate the relationship between the frequency bands of oscillations in the brain, Buzśaki and his colleagues defined ten different classes of frequencies (Figure 1) (Penttonen and Buzsáki, 2003; Buzsáki and Draguhn, 2004). Interestingly, a common ratio (equal to the Neper's number e) was identified between the mean frequencies of adjacent frequency bands resulting in a linear progression of mean frequencies on a logarithmic scale.

Among early studies inquiring the behavioral correlates of oscillations using EEG, slow oscillations including delta band oscillations were observed primarily during states of sleep or inactivity in humans as well as in animals (Norton and Jewett, 1965; Novak, Lepicovska and Dostalek, 1992; Achermann and Borbély, 1997; Amzica and Steriade, 1998). Theta band oscillations were detected in awake cats during orienting reflex (Grastyán *et al.*, 1959), which was quickly followed by several other studies that found associations between hippocampal theta and behaviors such as arousal, attention, movement, memory, visual search, decision-making, learning and sensorimotor processing (Buzsaki, 2006).

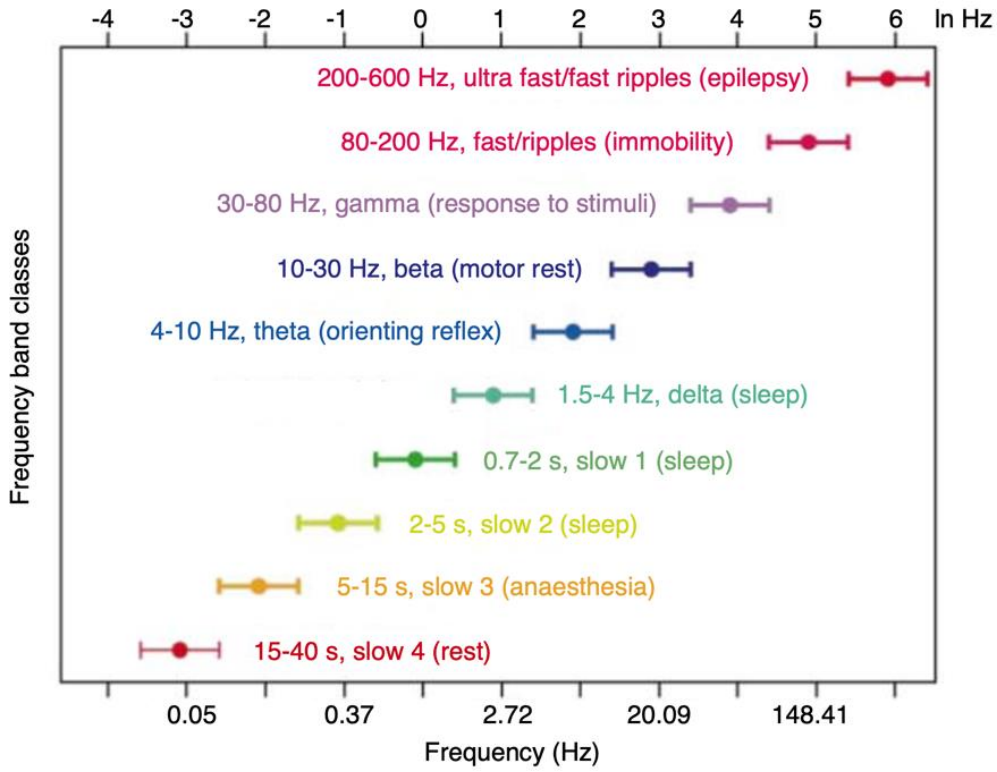

Figure 1: Categorization of oscillations in the brain into frequency band classes. For each class, its range of frequencies is shown along with its commonly used name. Note that the ranges defined here have not been strictly followed and across studies there is variability. An example behavioral correlate of each frequency band is included in brackets. Adapted from Buzsáki and Draguhn (2004).

Oscillations in the beta band have been associated with motor rest (Jasper and Penfield, 1949) and gamma band with response to visual stimuli (Regan, 1968). The works of Jasper and Penfield were also crucial to establishing the technique of

intracranial EEG (iEEG) or Electrocorticography (ECoG). An example ECoG recording made from the movement related areas of the brain illustrates beta oscillations during the relaxation period followed by a suppression of beta during movement execution (Figure 2). EEG and ECoG techniques, however, are not suitable to record higher frequency oscillations originating from deeper regions of the brain because of the frequency dependent attenuation of field potentials with distance (Logothetis, Kayser and Oeltermann, 2007). Microelectrodes, such as the tungsten microelectrode (Hubel, 1957) that can be inserted in the target areas of the brain, are employed for this purpose. Using microelectrodes inserted in the hippocampus of human patients (Bragin *et al.*, 1999) and animals (Buzsáki *et al.*, 1992) higher frequency oscillations covering the fast and ultra-fast frequency spectra have been recorded.

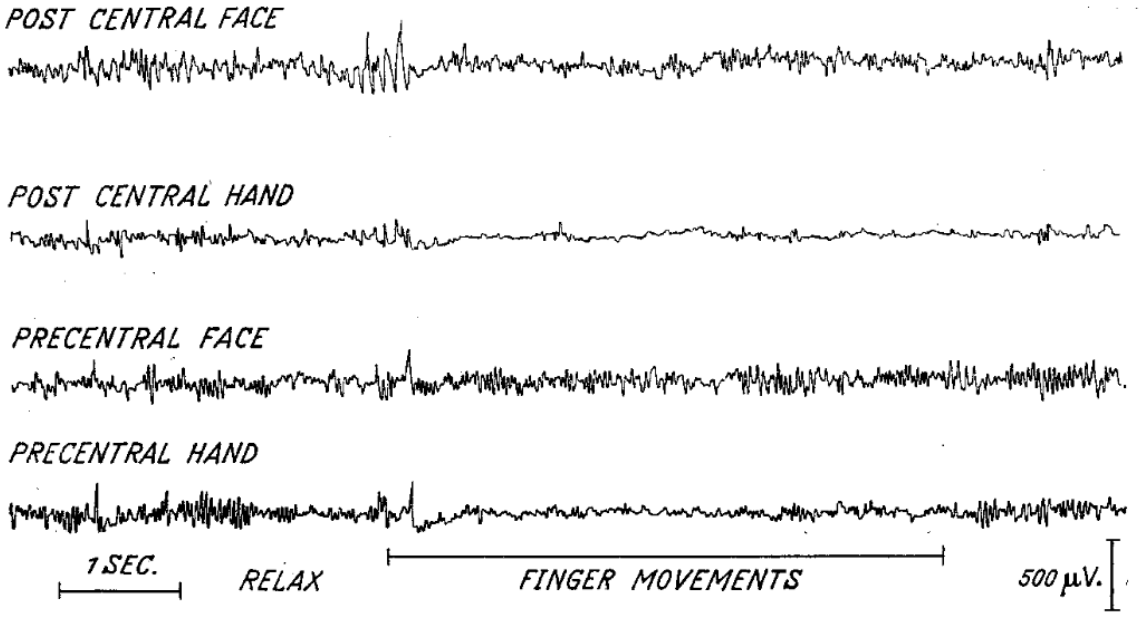

Figure 2: ECoG signals recording during a finger movement task. ECoG signals from post- and pre- central face and hand areas recorded during digit movements composed of successive touching of individual digits to the thumb. Adapted from Jasper and Penfield (1949).

1.1 Neuronal signal recording techniques

The activity of the brain can be recorded at three different scales: micro-scale, meso-scale and macro-scale. Micro-scale includes the level of individual neurons and

neuron compartments such as the cell body, dendrites, axons, ion channels, and synapses. Meso-scale includes the level of local populations of neurons (around tens of thousands), whereas macro-scale corresponds to the level of areas of the brain. Detailed analyses of signals recorded from each of these levels with appropriate spatial as well as temporal resolution are ideal to address how different levels of brain dynamics together produce function (Buzsáki and Christen, 2016). Although there is no one-fit-for-all solution, experimental and technological progress has made it possible to record neuronal signals with varying degrees of tradeoff between spatial and temporal resolutions (Sejnowski, Churchland and Movshon, 2014).

Another factor to consider when recording neuronal signals is the degree of invasiveness and the amount of neuronal damage induced by the recording technique. Functional magnetic resonance imaging (fMRI) is a non-invasive imaging technique in which neuronal activity is indirectly estimated based on the oxygenation level of the blood flowing near the neuronal structures (Logothetis *et al.*, 2001). Using fMRI, the 3D volume of the whole-brain can be captured at the meso- and macro-scales, albeit at very low sampling rates (typically below 1 Hz). Because of its indirect approach, fMRI would fail to capture potential coding schemes of the brain involving temporal correlations that are not associated with concurrent changes in firing rate (Buzsaki, 2006). Calcium imaging is another recording technique that has a similarly low sampling rate. However, it targets the micro-scale and has been employed to record simultaneously from almost all neurons of the brain in small animals such as the zebrafish (Ahrens *et al.*, 2013). Furthermore, by combining calcium imaging technique with electron microscopy, the functional properties of neurons such as its response to stimuli of different orientations can be studied alongside its structural properties such as its incoming and outgoing projections (Bock *et al.*, 2011). Although calcium imaging technique is very promising, it has largely been restricted to studies in smaller animal models. Its application in larger animals such as non-human primates faces a number of challenges because of the differences in the physical scale and the increased need for experimental longevity, which still needs re-engineering efforts (O'Shea *et al.*, 2017).

In comparison to the more recently developed optical technique discussed above, electrophysiological recordings date back to more than a century considering the electrical activity of the mammalian brain demonstrated by Richard Caton (Caton, 1875; Millett, 2001). Currently, there is a multitude of electrophysiological recording options that can record at submillisecond temporal resolution and at all three scales (micro, meso, and macro) of brain activity (Figure 3).

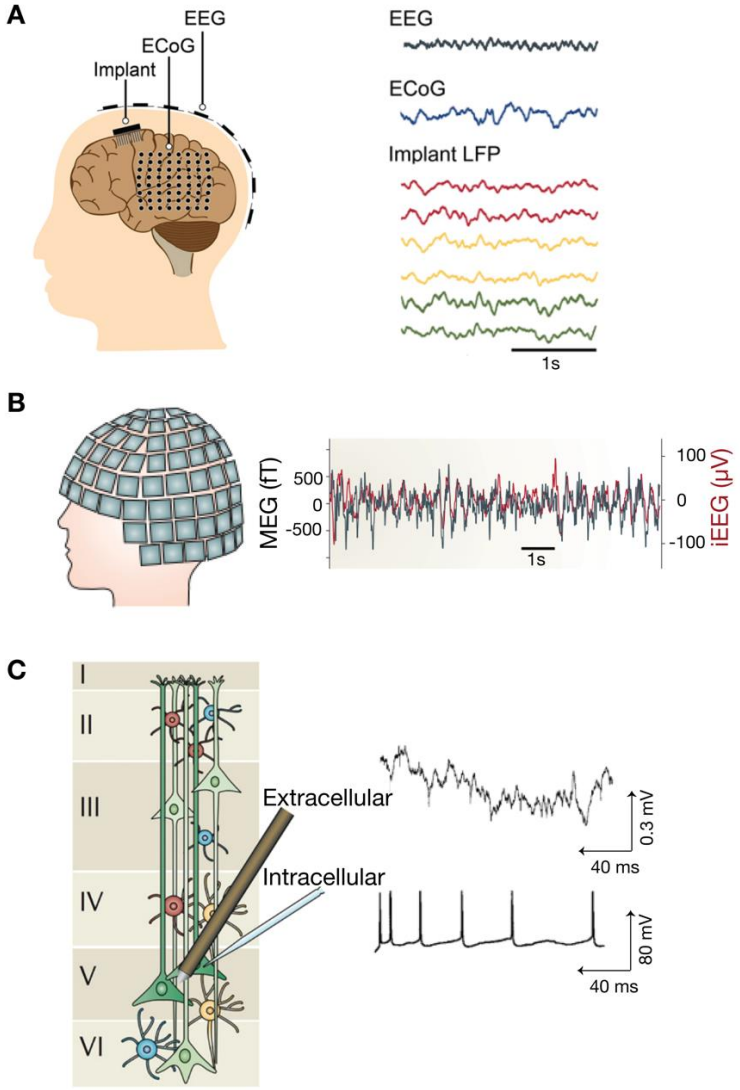

Figure 3: Illustration of electrophysiological recording techniques. A. Macroscopic recording using EEG and mesoscopic recording using ECoG and implantable microelectrodes (left). Example waveforms from each technique is shown to the right. **B.** MEG recording electrodes (left). Example MEG (black) and depth EEG (red) recordings. **C.** Neurons illustrated across cortical layers along with microelectrode probe (for extracellular recording) and glass pipette (for intracellular recording) (left). Example extracellular and intracellular recordings (right). Adapted from Buzsaki et al., (2012) and Obien et al., (2015).

At the micro-scale, the patch clamp technique (Neher and Sakmann, 1992) makes it possible to record intracellular currents at the level of single neurons (Figure 3C). However, this technique is mainly performed in-vitro on slices of the brain, with very few studies carried out in-vivo. The experiments conducted in animal models were limited to recording 2 - 4 neurons in parallel that were spatially less than 100 μm apart (Poulet and Petersen, 2008; Jouhanneau *et al.*, 2015).

At the macro- and meso-scales, a majority of electrophysiological techniques record field potential signals reflecting mainly the synchronous activity of hundreds to thousands of neurons (Buzsáki, Anastassiou and Koch, 2012). Examples of such electrophysiological recordings are EEG (when measured from outside the skull), magnetoencephalogram (MEG) (when magnetic field induced by this activity is recorded from outside the skull), ECoG (when measured using subdural electrodes placed on the cortical surface inside the skull), and local field potential (LFP) (when recorded using microelectrodes inserted in the brain) (Figure 3).

Tungsten microelectrode recordings pioneered by Hubel's work (Hubel, 1957) has inspired several generations of single and multi-electrode designs and recoding techniques (Hong and Lieber, 2019). Single-neuron recordings are performed by inserting a microelectrode in the brain and adjusting its position until neural activity is detected (Figure 3C). This method offers manual control on the recording site. However, it can lead to preferential selection of neurons that respond to a particular stimulus or task condition (Harris *et al.*, 2016). Such sampling bias is avoided by using microelectrode arrays with many recording sites that are chronically implanted in the brain. Michigan probe (Wise, Angell and Starr, 1970) and Utah array (Rousche and Normann, 1998) represent two distinct and widely employed multi-electrode array architectures (De Vittorio, Martiradonna and Assad, 2014). Michigan probes include distributed recording sites along the length of its silicon shanks. Neuropixels probes (Jun *et al.*, 2017) is a notable recent development in Michigan-type neural probes that tremendously increased the number of recording sites (n = 960) that can be selected for recording. In contrast to Michigan-type probes, Utah arrays (Rousche and Normann, 1998) consist of multiple rigid microelectrodes mounted on a

lightweight platform. Individual electrodes in these arrays are separated from each other by a few hundred micrometers, and the recording sites are present at the tip of each electrode.

Although architecturally similar to Utah arrays, floating microelectrode arrays (Figure 4A) were manufactured using a different fabrication technique that achieved over five-fold increase in electrode length (Musallam *et al.*, 2007). With this increase in the length of electrodes, deep sulci and other deeper neural areas became accessible. Furthermore, floating microelectrode arrays allow mixing electrodes of different lengths, electrode spacing, and electrode impedance, thereby extending the range of applications of microelectrode arrays. As the brain moves relative to the skull, these electrodes move freely with the brain, thereby minimizing damage to the surrounding tissue. The signals recorded from each electrode of the array consist of action potentials or spikes of neurons in the high frequency range (~300-5000 Hz) and LFPs in the low frequency range (~1-100 Hz) (Figure 4A).

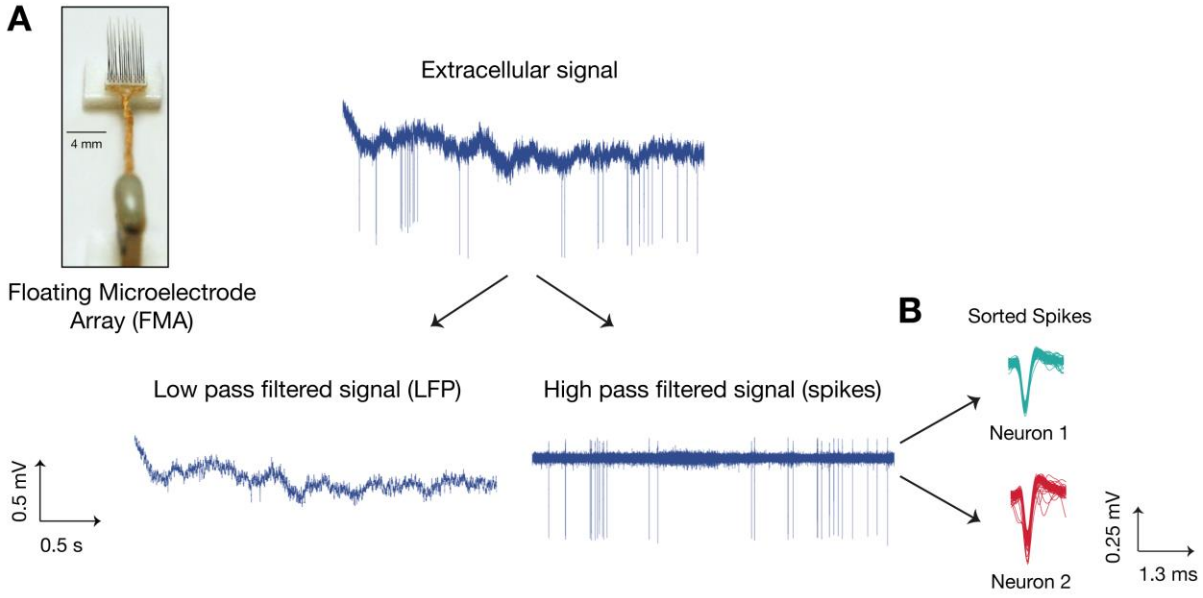

Figure 4: Exemplar LFP and spike signals recorded a microelectrode array. **A.** A 32 channel Floating microelectrode array (FMA) from Microprobe Inc (shown at a top-left corner). Example raw extracellular signal recorded from the FMA. Low pass filter is applied to the raw signal to obtain LFP signal; high pass filter is applied, followed by spike sorting to segregate spikes of individual neurons. **B.** Spike waveforms of two neurons isolated by spike sorting are illustrated. Adapted from Dann (2017).

Each electrode in the array records spikes from multiple neurons that are

present in the vicinity of its recording site. Spikes of neurons are considered to be all or nothing potentials, which can be isolated using spike-sorting algorithms (Quiroga, Nadasdy and Ben-Shaul, 2004; Dann *et al.*, 2016; Rossant *et al.*, 2016). Because the voltage deflection caused by a neuron is similar over all of its spike events, spike sorting algorithms exploit this similarity to identify and cluster spike waveforms with distinct features (Figure 4B), thus isolating neurons.

1.2 Functional connectivity

The brain is a dynamical system in which neuronal interactions are thought to be central to generate behavior, perception and cognition. Interactions between neuronal signals can be quantified using different measures of functional connectivity (Bastos and Schoffelen, 2016), which capture the temporal correlations in the signals. Measures of functional connectivity can be divided into directed and non-directed, which can be further sub-divided into model-based or model-free categories (Bastos and Schoffelen, 2016). Non-directed measures are designed to capture some form of signal interdependences, whereas directed measures seek to identify causal influences between signals. Model-based measures are developed assuming linearity in the signal interactions and are computationally more efficient than their model-free counterparts that can also capture non-linear interactions. A few exemplar measures of functional connectivity for the different categories are illustrated (Figure 5).

Rhythmic co-fluctuations in neuronal activity referred to as oscillatory synchrony is an important type of neuronal interaction that has been proposed to support flexible and effective communication (Fries, 2005, 2015). Oscillatory synchrony has been increasingly documented during many behaviors (Scherberger, Jarvis and Andersen, 2005; Sirota *et al.*, 2008; Gregoriou *et al.*, 2009; Canolty *et al.*, 2010; Bosman *et al.*, 2012). Coherence and phase locking value are the frequently used measures to capture oscillatory synchrony, which were also employed in the above studies. Both measures take values ranging from 0 to 1, with 0

corresponding to not at all synchronized and 1 corresponding to fully synchronized signals. The first step to compute coherence and phase locking value is to transform neuronal signals from the time domain to frequency domain using techniques such as Fourier transformation. Following this, coherence is calculated by taking the absolute value of the cross spectrum of the two signals, which is then normalized by the geometric mean of their individual auto-spectra (Mitra and Pesaran, 1999). The computation of phase locking value is similar to coherence, however, it includes a minor modification: signals in the frequency domain are amplitude normalized prior to computing cross spectrum. Thus, phase locking value (Lachaux *et al.*, 1999) reflects phase synchronization more strictly than coherence and mitigates amplitude correlations from affecting the measured phase locking values (Bastos and Schoffelen, 2016).

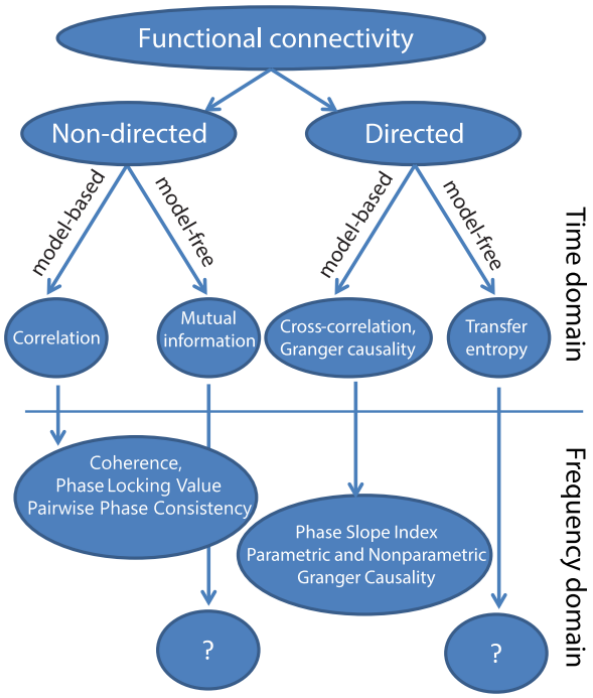

Figure 5: Taxonomy of functional connectivity metrics. Measures of functional connectivity are divided into directed and non-directed, which is further subdivided into model-based or model-free categories in the time domain. When a frequency domain adaptation of a method exists, it follows the same categorization as its time domain counterpart. Adapted from Bastos and Schoffelen (2016)

The type of recorded signal is another crucial factor to consider in the analysis of functional connectivity. Three different types of measures of functional

connectivity exist based on the type of neuronal signal: field-field, spike-spike, and spike-field measures (Figure 6). While only pairwise interactions are depicted in the illustration (Figure 6), given a dataset with a large number of simultaneously acquired signals, functional connectivity can be measured between all pairs of recorded signals to obtain networks of functional connectivity (Dann *et al.*, 2016; Nigam *et al.*, 2016).

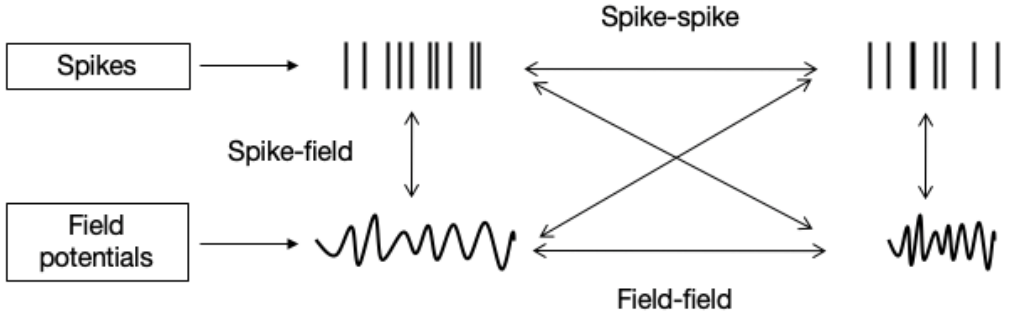

Figure 6: Synchrony between different types of signals. Illustration of binary spike trains and continuous field potential signals and the three different types of interaction between them. Figure inspired from Arce-McShane et al., (2016)

1.2.1 Field-field synchrony

Field-field coherence is largely employed to examine neuronal synchrony (Brovelli *et al.*, 2004; Bosman *et al.*, 2012; Nácher *et al.*, 2013; Bastos *et al.*, 2015) because recording field potentials is relatively easier (see section 1.1). Furthermore, the measures of functional connectivity, such as coherence and phase locking value described in the previous section are directly applicable to field potentials, whereas some modifications are required to apply them to binary spike signals.

Enhanced field-field coherence in the beta band (14 – 30 Hz) has been observed during a motor task in which monkeys pressed and held a hand lever (Brovelli *et al.*, 2004). In addition to computing coherence, Granger causality (Dhamala, Rangarajan and Ding, 2008; Seth, 2010) was also analyzed by Brovelli and colleagues revealing directed functional connectivity from parietal areas to central and post-central motor areas. Delta (1 – 4 Hz) coherence has also been reported in the fronto-parietal areas during a somatosensory discrimination task suggesting low frequency synchrony to

be involved in decision-making task processes (Nácher *et al.*, 2013). In contrast to fronto-parietal movement related areas discussed above, gamma coherence is predominantly documented in the visual areas of the brain. Modulation of gamma (60-80 Hz) coherence with attention has been demonstrated between visual areas V1 and V4 during a selective attention task (Bosman *et al.*, 2012), suggesting the selection of relevant input for further processing by increase gamma synchrony. To examine the role of the different frequencies of inter-areal coherence, field potential signals were recorded in parallel from 8 different brain areas (Bastos *et al.*, 2015). Then, coherence and Granger causality analyses were conducted, which revealed theta (~ 4 Hz) and gamma (60-80 Hz) synchrony to carry feedforward and beta (14-18 Hz) synchrony to carry feedback influences.

It is important to note that due to volume conduction, field potential signals spread via the neural tissue across large cortical volumes, thereby confounding the measured field-field coherence values. To address this problem, methods such as weighted phase lag index (Vinck *et al.*, 2011) and the imaginary part of coherency (Nolte *et al.*, 2004) have been developed. These methods exploit the fact that volume conduction predominantly affects the real part of cross-spectral density because the cortical tissue has been demonstrated to be resistive and not capacitive (Logothetis, Kayser and Oeltermann, 2007). By ignoring the real part and employing only the imaginary part of the cross-spectral density for coherence analysis, volume conduction effect is eliminated.

However, volume conduction is not the only problem affecting field-field measures. Recently, it has been demonstrated that field-field measures such as coherence as well as Granger causality are confounded by the synaptic mixing of local and afferent inputs in the LFP signal (Pesaran *et al.*, 2018; Schneider *et al.*, 2020). To elaborate, spikes in an area contribute to synaptic potentials not only in the same area but also in other anatomically connected areas. This results in correlated components in the field potential signals that are recorded from any two anatomically connected areas, which are in turn captured by field-field measures of coherences. Therefore, even in the absence of true oscillatory coupling between two areas, higher

values of coherence can be obtained due to synaptic mixing. Unlike volume conduction, synaptic mixing is not instantaneous therefore, it cannot be removed using the solution developed for volume conduction (Schneider *et al.*, 2020) that was discussed in the previous paragraph. Given the synaptic mixing problem, strong entrainment of neurons to the oscillatory rhythm has been proposed as a prerequisite to establish functionally relevant coupling between two distinct oscillators (Schneider *et al.*, 2020).

1.2.2 Spike-spike and spike-field synchrony

Oscillatory coupling of spikes can be quantified using spike-spike and spike-field measures. Cross correlation histogram (CCH) obtained by calculating Pearson correlation coefficient at incremental time lags applied to one neuronal signal with respect to the other is a commonly used measure that is employed to capture spike-spike synchrony (Gray *et al.*, 1989; König, Engel and Singer, 1995; Dann *et al.*, 2016). In cat V1, gamma (40-60 Hz) synchronization was discovered by researchers examining spike-spike interactions (Gray *et al.*, 1989). V1 neurons in cats exhibit not only short (within hemisphere) but also long-range (across hemisphere) gamma (30-70 Hz) synchrony during passive viewing of visual stimuli (König, Engel and Singer, 1995). More recently, low (3-7 Hz) and beta (18-35 Hz) synchronized neurons were found in fronto-parietal areas during a delayed grasping task that included a decision component (Dann *et al.*, 2016).

Spike-field measures capture how neurons preferentially fire spikes at particular phases of the LFP signal (Pesaran *et al.*, 2018). Spike-field measures have several advantages in comparison to field-field and spike-spike measures. First, oscillatory synchrony can be analyzed with single neuron resolution using spike-field measures of coherence. This mitigates the problem of synaptic mixing, which was earlier discussed in the context of field-field measures. While spike-spike measures also satisfy this requirement, they suffer from an inherent rate bias due to the masking of subthreshold potentials by the spiking threshold of neurons (Cohen and Kohn, 2011). Also, spike-spike interactions reflect a small subset of all network

interactions due to the restricted number of neurons that can be recorded simultaneously using electrophysiology. Second, spikes and LFPs carry different and complementary types of information: spikes are the outputs generated by neurons, whereas LFPs reflect the inputs and local processing (Scherberger, Jarvis and Andersen, 2005; Pesaran, 2010). By including these complementary signals in the estimation of oscillatory synchrony, a larger part of the underlying network is sampled.

Coherence (Mitra and Pesaran, 1999) is a frequently used measure to quantify spike-field synchrony (Pesaran *et al.*, 2002; Brovelli *et al.*, 2004), although it has a known dependency on spike rate (Lepage, Kramer and Eden, 2011). This dependency makes it difficult to use this measure to compare neurons and experimental conditions with different spike rates as it requires further corrective steps. A method based on generalized linear modeling was proposed to overcome the problem of rate-dependence of coherence computed between spikes and LFPs (Lepage *et al.*, 2013). However, this solution is model-dependent and therefore has to be carefully interpreted (Aoi *et al.*, 2015). Other corrective measures include spike-thinning procedures (Gregoriou *et al.*, 2009), but they can result in the removal of a considerably large number of spikes from neurons, especially when the firing-rate differences between experimental conditions are large. Furthermore, this approach of randomly removing spikes until the mean firing rate matches assumes Poisson or non-history dependent firing neurons, which is often violated in real neurons (Aoi *et al.*, 2015).

Phase Locking Value (PLV) (Lachaux *et al.*, 1999) is another commonly used spike-field measure that quantifies the homogeneity in the LFP phases at which the neuron fires. However, for small sample sizes PLV has been shown to overestimate synchrony (Vinck *et al.*, 2010). To overcome this problem, pairwise phase consistency (PPC), a novel measure of spike-field phase locking was introduced (Vinck *et al.*, 2010, 2012). PPC implementation (Figure 7) removes rate bias by introducing a pairwise approach to computing phase similarity. Furthermore, bias due to history effects such as bursting are also be removed by considering pairs of spike-LFP phases recorded

from different trials during PPC calculation. PPC is calculated by taking the average of the cosine transformed phase differences between all pairs (except the within trial pairs) of spike phases measured relative to the LFP signal. PPC value of 1 corresponds to perfect synchrony and 0 corresponds to the absence of synchrony. However, as a consequence of being an unbiased measure, PPC has higher variance at lower firing rates and can sometimes take negative values (Vinck *et al.*, 2010). To address this issue, studies employing PPC often set a fixed number of spikes (e.g., 50 spikes) as the minimum number of spikes required for PPC computation (Vinck *et al.*, 2013; Onorato *et al.*, 2020). It is also important to note that LFP signals especially at higher frequencies are contaminated by spikes recorded from the same channel (Waldert, Lemon and Kraskov, 2013). Thereby, typically spikes and LFPs recorded from different electrodes that are separated by several hundred micrometers are used in spike-field coherence analysis.

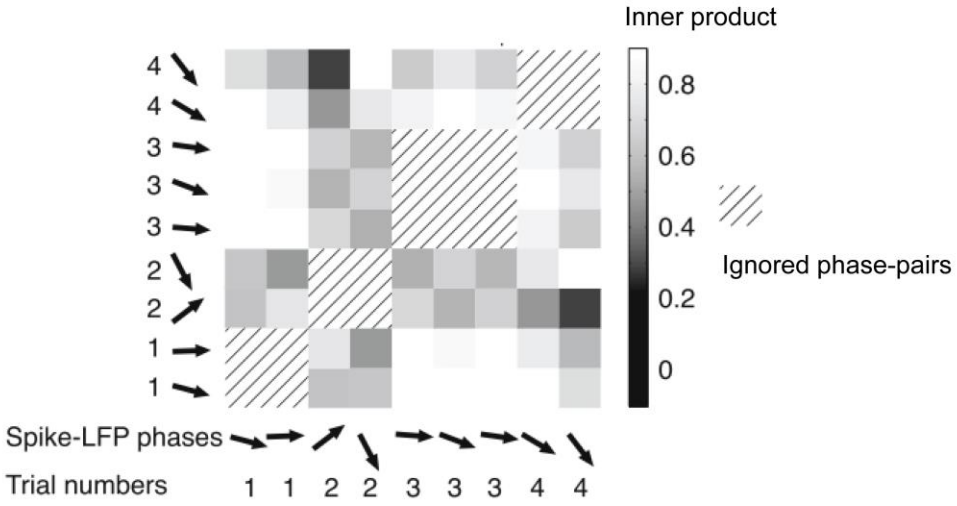

Figure 7: Illustration of pairwise phase consistency. The phases of the spikes measured relative to the LFP signal called spike-LFP phases are shown along the left and bottom margins. The numbers included with each phase vector indicates the trial from which the spikes were recorded. Gray scales in the matrix indicate the similarity between the spike-LFP phase pairs shown along the corresponding row and column. PPC is calculated by averaging over all the dot products (excluding the within trial spike-LFP phase pairs). PPC illustration adapted from Vinck *et al.*, (2012)

Given these advantages, the number of studies using spike-field measures is growing. In agreement with visual system studies discussed earlier, gamma oscillatory synchrony has also been captured between neurons and field potential

signals using PPC in V1 across mice, capuchin and macaque monkeys during fixation tasks (Onorato *et al.*, 2020). Similarly, in monkey V4, neurons whose receptive field overlapped with the attended stimuli exhibited increased gamma synchrony with local neuronal populations (Fries, 2001). Motor system studies that included recordings from the parietal cortex have predominantly documented beta spike-field coherence during movement planning. For example, in a center-out reach task, enhanced synchrony in the beta band was observed during reach planning (Scherberger, Jarvis and Andersen, 2005). Recently, more evidence has been accumulated supporting the role of parietal beta synchrony in movement preparation (Pesaran, Nelson and Andersen, 2008; Dean, Hagan and Pesaran, 2012; Wong *et al.*, 2016). Since beta synchrony across studies has been mainly observed during steady states of the task, i.e., when neither the external stimulus nor the behavior changes, it has been inferred that beta synchrony has an important role in maintaining steady state (Engel and Fries, 2010).

A detailed tabulation of many other studies including their task, recording details, and the documented behavioral correlates of spike-field oscillatory synchrony is included (Appendix B : supplementary tables) separately for motor (Table 1), visual (Table 2) and other cognitive tasks (Table 3).

1.3 Selective and flexible information processing

How the collective activity of neurons underlies different functions of the brain such as perception, cognition and behavior is still not fully understood. Several theories has been put forth to address this question. Donald Hebb in his seminal work titled "The Organization of Behavior" proposed that the ability of the brain to generate coherent thought stems from the activation of many neurons organized in distinct groups (Hebb, 1949). Describing how some neuron pairs might form stronger connections, Hebb said, "*When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing*

B, is increased." Extending this idea to many neurons, he coined the term "cell assembly" to describe groups of neurons presumably self-organized by synaptic strengthening caused by contiguous firing. Sequential activation of different cell assemblies was proposed as the neuronal substrate of perception and other cognitive capabilities such as planning, decision making, learning, and memory.

While the theory of cell assembly was proposed in 1949, experimental evidence supporting this hypothesis took much longer due to limitations on recording large number of neurons simultaneously. Recordings of hippocampal pyramidal neurons (up to 68 neurons recorded simultaneously) provided the basis to test the hypothesis on cell assemblies (Harris *et al.*, 2003). Cell assembly hypothesis predicts that the firing rate of neurons includes two different components. First, the part that is modulated by external sensory input; second, the part that is internally generated and shared between neurons belonging to the assembly. As a consequence, firing rate of neurons can be better predicted from the combination of recorded activity of other neurons which belong to the same cell assembly and the external behavioral information. In agreement with this, the prediction of spiking of individual neurons was demonstrated to improve when spiking of other neurons was employed in addition to the information derived from behavior. Furthermore, by using a time window based approach to the prediction process, the time windows within which the spike times of individual neurons were best predicted from the population was shown to vary between 10ms and 30ms, suggesting that the neurons are organized in assemblies whose activity synchronizes transiently. This time-scale might be optimal for information processing as it is within the time required for the post-synaptic potentials arriving at the hippocampal pyramidal neurons to decay, thereby the spikes from the cell assembly arriving within this time window can be effectively integrated by the receiving neuron (Harris *et al.*, 2003). Intriguingly, this time window also corresponds to the time period of gamma oscillations suggesting gamma oscillatory synchronization of the cell assembly.

A limitation of the cell assemblies concept introduced by Hebb is that when two different cell assemblies are simultaneously activated, there is no explicit

mechanism by which they can be distinguished (von der Malsburg, 1981). For example, when multiple objects are present in the visual scene, each object activates a different group of neurons in the brain. When the objects in the visual scene are overlapping, their representation in the primary visual cortex would also be overlapping due to the retinotopic organization of neurons, which means that the neurons are anatomically arranged such that they reflect the spatial relationship between the stimuli that activated them. To disentangle this, the cell assemblies approach was extended under the binding by synchrony (BBS) hypothesis. BBS was put forth to address the multiple-object encoding problem in the visual system (von der Malsburg and Schneider, 1986; Singer, 1993; Singer and Gray, 1995). BBS predicted that neurons activated by different objects are synchronized to different oscillatory rhythms. Thereby, multiple cell assemblies can be simultaneously active while still being distinguishable. An implicit goal of the BBS theory was to explain how object attributes such as colors and contours are bound together in its neuronal representation. In a contour grouping task carried out in monkeys this theory was tested on data recorded from area V1 (Roelfsema, Lamme and Spekreijse, 2004). In this task, firing rate of neurons was found to correlated to the grouping of contours, whereas synchrony was found to be uncorrelated, (Roelfsema, Lamme and Spekreijse, 2004), thereby disproving BBS.

The theories discussed above focused on how oscillatory synchrony might coordinate spatially distributed neuronal activity into a representational code. However, theory on how flexible neuronal communication might be facilitated by oscillatory synchrony has not yet been discussed. Before going into the depth of this topic, a few definitions crucial for the following discussion that have been adopted from literature (Fries, 2015) are presented. Firstly, neuronal representation is defined as the 'spatial activation pattern in a group of neurons'. Secondly, neuronal communication is defined as the 'transfer of one representation in the presynaptic or sending group to a new representation in a postsynaptic or receiving group'. Lastly, neuronal computation is the 'transformation that happens between the representations'. With these definitions in place, communication can be examined as

an important computational process that transforms inputs into new representations in the brain, which sometimes leads to behavior. Traditionally, neuronal communication is associated with anatomical connectivity. However, sometimes different activation patterns of neurons is elicited by the same external stimulus due to internal factors such as attention, past experiences, and internal choice. In visual areas, gamma synchrony has been demonstrated to be modulated by attention (Fries, 2001; Gregoriou *et al.*, 2009). Communication through coherence (CTC) proposes that oscillatory neuronal synchronization facilitates the formation of flexible communication structures supporting cognitive functions such as selective attention (Fries, 2005, 2015).

To elaborate, during the gamma cycle it is known that excitatory neurons are first activated leading to a window of opportunity for incoming signals to be integrated (Buzsáki and Wang, 2012; Fries, 2015). This time window is very short and lasts for only about 3ms, soon after the local inhibitory neurons are activated. The next cycle of gamma oscillation can only begin after the activity of the inhibitory neurons decays. To be effective, the inputs from the pre-synaptic neuronal group have be timed such that they arrive at the post-synaptic neuronal group during its excitation phase in each cycle of gamma oscillation (Figure 8). Given the repetitive nature of oscillatory signals, the theory of communication through coherence (CTC) proposed that oscillatory synchrony subserves the precise synchronization between the pre- and post-synaptic groups of neurons (Fries, 2015). Furthermore, dynamic changes in neuronal coherence can lead to selective processing of behaviorally relevant information in a hard-wired anatomical neuronal network. Experimental evidence for selective communication, as hypothesized by CTC, has been obtained from primates performing a selective attention task in which they were cued to attend one of two visual stimuli presented on a monitor (Bosman *et al.*, 2012). In this study (Bosman *et al.*, 2012), evaluation of field-field coherence revealed that the V1 neurons activated by the attended stimulus were preferentially synchronized with V4 neurons while V1 neurons activated by the unattended stimulus were not synchronized. From this finding, selective communication in the brain has been inferred to be facilitated

by oscillatory synchrony. However, it was recently demonstrated that inter-areal field-field coherence can be precisely predicted by anatomical connectivity and oscillatory synchronization in the sending area (Schneider *et al.*, 2020). These findings argue against the theory of communication through coherence and suggest that it might be the other way around, that is, coherence through anatomical connectivity, and hence communication.

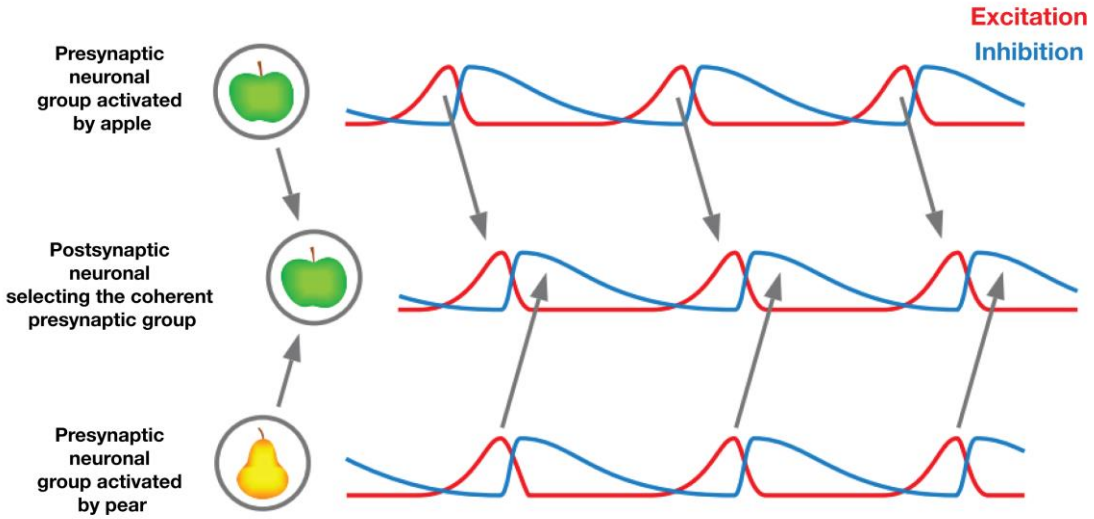

Figure 8: Selection of information from lower to higher visual stream via CTC. Each group undergoes a network-level excitation phase (around the peak of red traces), which is followed by an inhibition phase (around the peak of blue traces). As the input from the presynaptic neuronal group that represents the apple reaches the post-synaptic neuronal group during its excitation phase, this information is optimally transmitted, whereas the representation of the pear is suppressed. Adapted from Fries (2015).

Alternatively, another theory was proposed describing the syntax or grammar of neural communication (Buzsáki, 2010). This theory provides the following objective, and reader-centric definition of cell assemblies: neurons that fire within the integration time window of the downstream reader define a cell assembly (Buzsáki, 2010). Here, it is noteworthy that neuronal synchrony, which is central to the formation of cell assemblies, can only be objectively defined from the perspective of a downstream neuron or more generally a downstream reader/observer and not from the stimulus based bottom-up approach used earlier (e.g. in BBS). Furthermore, in the modified definition of cell assembly the functional effect of the assembly on the reader is emphasized, whereas the requirement on the synaptic connectivity of the

members of the cell assembly is relaxed. The simplest example of a reader is a neuron and the functional effect the cell assembly has on the reader is binary: action potential or no action potential. The reader in this model need not necessarily be a neuron, it can also be a cell assembly that can further be combined by the next set of downstream readers to generate neural words. There are numerous advantages to cell assembly-based implementation of functions of the brain. First, single neurons can be vulnerable to potential failures in spike transmission, whereas an assembly has higher failure tolerance because for the reader spike rate variations in individual neurons do not matter as much as the collective intensity of the activity of the assembly. Second, nearly unlimited cell assemblies are possible thereby greatly enlarging number of possible representations that can be encoded in the brain (Buzsáki, 2010).

Empirical evidence supporting the formation of cell assemblies has been largely obtained from studies on hippocampal cells involved in navigational tasks. LFP signals recorded from the hippocampus oscillating in the theta frequency band for instance have been demonstrated to phase modulate neocortical cell assemblies of neurons oscillating in gamma frequency band (Sirota *et al.*, 2008). In this case the cell assembly formation was facilitated by gamma oscillations whereas the theta oscillations in the hippocampal LFPs provided a mesoscopic reader mechanism that temporally sequenced the cell assemblies. Importantly, in this framework as well, oscillations in the brain have been suggested as a potential mechanism for the coordination of cell assemblies. Taken together, oscillations have been commonly proposed across the above discussed theories as the central mechanism for selective communication across spatially distributed areas in the brain.

Modelling studies simulating networks of connected single neurons have also demonstrated how transient oscillations of the brain can flexibly modulate information flow, which is essential for context-dependent behavior (Palmigiano *et al.*, 2017). However, empirical evidence on network reconfiguration is available predominantly at the level of macro- and meso-scales from studies examining changes in whole brain networks over time (Leonardi *et al.*, 2013; Allen *et al.*, 2014;

Khambhati, Mattar and Bassett, 2017), behavioral conditions (Cole *et al.*, 2014; Hearne *et al.*, 2017; Thompson, Brantefors and Fransson, 2017) including pathological conditions (Bassett *et al.*, 2009). To improve the biological plausibility of the simulated models the dynamics of oscillatory network structure must be characterized at the level of single neurons during behavior. Study of networks at the level of single neurons is scarce and has mostly been done in-vitro with the exception of a few experiments conducted in-vivo (Dann *et al.*, 2016; Nigam *et al.*, 2016). In these studies, the characterization of the single neuron resolved network structure revealed that the strongly connected neurons in the network form dense connections with one another. The resulting rich-club network topology has been linked to efficient communication of information (van den Heuvel and Sporns, 2013). Furthermore, the rich club was found to be predominantly composed of oscillatory neurons, which strongly synchronized within and between areas and may be central to flexible information processing (Dann *et al.*, 2016). However, how changes in the oscillatory network structure might support selective information processing that is essential for context-dependent behavior is largely unexplored particularly during sensorimotor transformations.

1.4 The fronto-parietal grasping network

Grasping behavior is an important function of the brain that enables direct interaction with the environment. These interactions include context-dependent behaviors requiring flexible selection and transformation of relevant sensory information into appropriate movement relevant activity. Cortical networks have to perform multiple complex computations even for simple hand-object interactions. The seemingly trivial task of grasping a coffee mug is carried out by executing a sequence of processing steps including perceptual, motor processes as well as sensorimotor transformations.

For visually guided grasping movements, visual assessment is the first step that informs about the physical properties of the mug such as its shape and size, as

well as its spatial properties such as location and orientation. At this stage, the perceived visual information is still in an eye-centered frame of reference and is transformed into a body-centered frame of reference for movement planning and execution (Batista *et al.*, 1999; Lehmann and Scherberger, 2013; Michaels and Scherberger, 2018). Then, depending on the context (e.g. whether to drink coffee or to move the mug, whether the mug is full or empty, cold or hot) a grip type is selected (Baumann, Fluet and Scherberger, 2009; Fluet, Baumann and Scherberger, 2010; Dann *et al.*, 2016). The selected grip type is executed by activating appropriate arm and hand muscles to reach and grasp the mug. As movement is being performed, sensory receptors in the digits provide feedback (e.g. on the grip-force being applied, texture) and any errors that might arise during the movement are monitored to apply corrective measures.

Visual processes underlying behaviors such as the one described above are carried out in two distinct pathways: the ventral and the dorsal pathway. Anatomically, both the pathways begin at the primary visual cortex (V1) and diverges with the ventral stream traversing the occipitotemporal cortex and the dorsal stream going through the occipitoparietal cortex (Figure 9). Functionally, the ventral and dorsal pathways are hypothesized to be involved in object identification and object-oriented action processes, respectively (Goodale and Milner, 1992). Patient case studies in which either the ventral or the dorsal pathway was damaged demonstrated the complementary roles played by these two pathways (Goodale *et al.*, 1991, 1994). In these case studies, two patients performed a shape discrimination and a grasping task. One patient (RV) who had sustained bilateral lesion along the dorsal stream was unable to correctly pre-shape her hand to perform object-directed movements, although she could correctly identify the objects. In contrast, the patient (DF) with a bilateral lesion in the ventral stream was unable to recognize objects although she was capable of performing grasping movements with appropriate grip aperture and orientation. These findings taken together with many other supportive evidence gathered over the last few decades suggest that the dorsal stream extracts the information from vision that is required for action, whereas the ventral stream

extracts the information from vision that is required for perception (Goodale, 2014).

Distinct parietal and frontal areas along the dorsal stream have been identified to be crucial for the planning and execution of different types of movements such as grasping, reaching and saccade movements (Luppino and Rizzolatti, 2000; Rizzolatti and Luppino, 2001). Ventral premotor area F5 in the frontal lobe and anterior intraparietal area AIP in the parietal lobe are strongly involved in grasping movements (Rizzolatti and Luppino, 2001; Janssen and Scherberger, 2015; Borra *et al.*, 2017) and are also known to be strongly interconnected (Luppino *et al.*, 1999). Inactivating either one of these areas leads to deficits in hand movements. The deficit in grasping following the inactivation of AIP was characterized by a mismatch between object properties and the hand shape used to grasp it (Gallese *et al.*, 1994). Similar deficits were also observed following the inactivation of F5 (Fogassi *et al.*, 2001). Further supporting the role of areas F5 and AIP in grasping, neurons in these areas have been found to respond strongly during the planning and execution of grasping movements towards specific objects (Rizzolatti *et al.*, 1988; Taira *et al.*, 1990; Murata *et al.*, 1997, 2000). Interestingly, Murata and colleagues (Murata *et al.*, 1997, 2000) also documented that some of the neurons that were strongly active during the grasping tasks were also similarly activated during fixation when the same objects were presented.

To dissociate grip- from vision-related neural coding, a physiologically plausible model of the fronto-parietal grasping network including areas F5 and AIP called the FARS (Fagg/Arbib/Rizzolatti/Sakata) model (Fagg and Arbib, 1998) was developed. A number of simulation experiments were carried out with this model. For the first simulation, objects with different visual appearance but grasped the same way were chosen; second, two different grasps were instructed for the same object with some trials having a delay between the go signal and the grip instruction; third, boxes of different heights and widths were grasped. Results from these simulations led to a number of hypotheses and predictions on the functional role of the F5-AIP grasping network, which is summarized in the illustration (Figure 9). Briefly, the model suggested that all potential action opportunities associated with an object

referred to as affordances are extracted from the visual features of the object in area AIP. This process is likely to have object identity information from the reciprocal connections existing between AIP and inferior temporal (IT) cortex of the ventral stream (Webster, Bachevalier and Ungerleider, 1995). Furthermore, AIP was also implicated in maintaining the affordances of the object in active memory, which is partly shared with F5 due to the strong anatomical connectivity between the two areas. F5, on the other hand, was hypothesized to select a grasp plan depending on factors such as the intention of the individual, and action goals. The key ideas introduced by the FARS models discussed above have been employed across many other models describing the cortical grasping network (Rizzolatti and Luppino, 2001; Cisek, 2007).

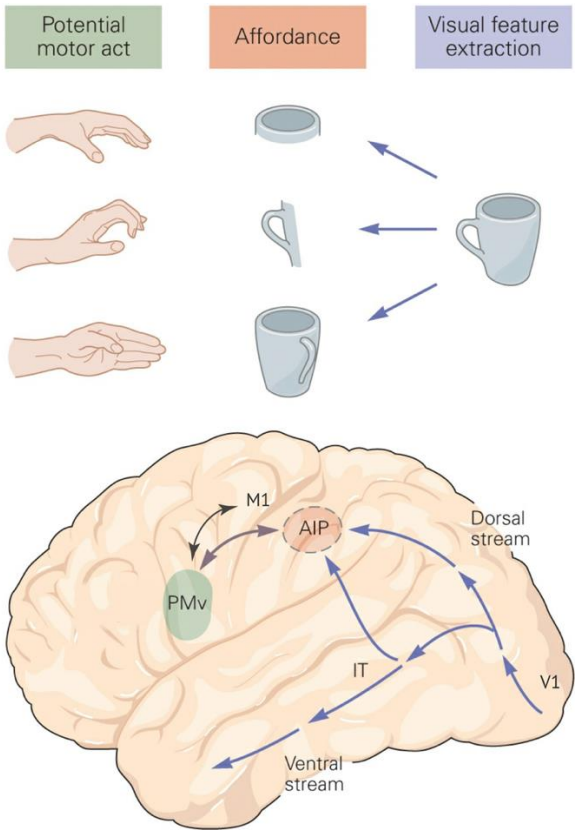

Figure 9: Sensorimotor integration in the fronto-parietal network during grasping. Areas of the brain involved in grasping and the computations carried out by them are at the bottom and top rows, respectively. Affordances, illustrated in the mid-row are the opportunities for action offered by the object. Figure modified from Kandel *et al.*, (2000).

Decoding analyses carried out on spikes from neurons recorded simultaneously from primates during a delayed grasping task (42-48 objects), lends strong support to the predictions of this model (Schaffelhofer and Scherberger, 2016). In their study, F5 neurons were found to be predominantly modulated by grip type whereas AIP neurons were found to be predominantly modulated by the visual object properties type. Interestingly, AIP neural population could discriminate differently shaped objects including abstract objects that were grasped similarly even when grasps were executed in dark, suggesting coding of not only visual features but also object affordances (Schaffelhofer and Scherberger, 2016). More evidence for coding of affordances in AIP was obtained from another delayed grasping task in which the objects were first presented and the grip type (power or precision) instruction was given after a delay (Baumann, Fluet and Scherberger, 2009). Neurons recorded from AIP coded both grip types simultaneously until the grip type instruction was provided.

Despite the supporting empirical evidence, a disadvantage of the earlier models of the grasping network is that they required manual tuning of neurons to fit the hypothesized roles of the areas, which might introduce subjective biases. Recently, an artificial neural network based model for the entire processing pipeline from the processing of visual inputs to the generation of grasp movements has been proposed (Michaels *et al.*, 2020). In addition to areas F5 and AIP, this model also includes primary motor cortex (M1), which is reciprocally connected to F5, and directly modulates motoneurons innervating the arm and hand muscles (Rathelot and Strick, 2009). The neurons of the model demonstrated visual- and movement-related information being present in all three areas, with a visuomotor gradient from AIP to F5 to M1 during the transformation of vision to action (Michaels *et al.*, 2020). In comparison to the earlier models, networks trained by Michaels and colleagues captured more variance in the recorded neural activity and also mirrored some of the inter-areal relationships in the neural data without being explicitly trained to do the same. Furthermore, some of the known deficits from lesions studies were also reproduced by the network, there providing a more biologically plausible model for

the generation of grasping movements.

Interestingly, low frequency oscillations were observed in the dynamics of recurrent neural networks (RNNs) during the execution of hand movements (Michaels *et al.*, 2020). Similar to this observation, RNNs trained on reaching movements also exhibited low frequency oscillations (Sussillo *et al.*, 2015; Michaels, Dann and Scherberger, 2016) during movement execution. While these RNNs were trained on smoothed spike signals that were further averaged over trials, networks trained on non-smoothed spike signals that were trial separated exhibited oscillations in the higher frequencies (15-40 Hz) during movement preparation (Pandarinath *et al.*, 2018). These findings strongly agree with the findings of low and beta frequency oscillations observed in the fronto-parietal areas related to grasping (Scherberger, Jarvis and Andersen, 2005; Dann *et al.*, 2016) and reaching (Martínez-Vázquez and Gail, 2018) movements. Importantly, the networks strongly matched the recorded neuronal data only when they were optimized to be simple, which resulted in the underlying solution being a low dimensional oscillator (Sussillo *et al.*, 2015; Vyas *et al.*, 2020). These oscillations might be facilitating effective and flexible communication within and between cortical areas (Fries, 2005, 2015) for sensorimotor transformations.

1.5 Tracking hand kinematics

Hands provide the means by which one can grasp and manipulate objects, thereby interacting with the environment. Tracking hand kinematics, which includes movement parameters such as position, and orientation is important not only to understand the motor behavior of the hand but also to investigate the neuronal circuitry generating hand movements. However, hand movement tracking poses challenges due to its remarkable dexterity observed particularly in primates. The complex hand movements result from the interaction between 27 bones and over 30 muscle-tendon units (Schwarz and Taylor, 1955; Jones and Lederman, 2006). The index, middle, ring and little finger of the hand each have 3 bones (proximal, middle

and distal phalanges), whereas the thumb has 2 bones (proximal and distal phalanges). The carpal and metacarpal bones are distributed in the palm and wrist (Figure 10A). Furthermore, the 27 degrees of freedom (DOFs) targeted to estimate hand pose and position (Figure 10B) are as follows. The digits of the hand, excluding the thumb, each have 4 DOFs: extension/flexion at the proximal interphalangeal (PIP), extension/flexion at the distal interphalangeal (DIP), extension/flexion and abduction/adduction at the metacarpophalangeal (MCP) joints. The thumb exhibits 5 DOFs: extension/flexion at the interphalangeal (IP), extension/flexion and abduction/adduction at the MCP and Carpometacarpal (CMC) joints. The wrist has 3 DOFs: extension/flexion, pronation/supination, and ulnar/radial deviation. Lastly, the hand is transported by the arm along three orthogonal directions leading to an additional 3 DOFs.

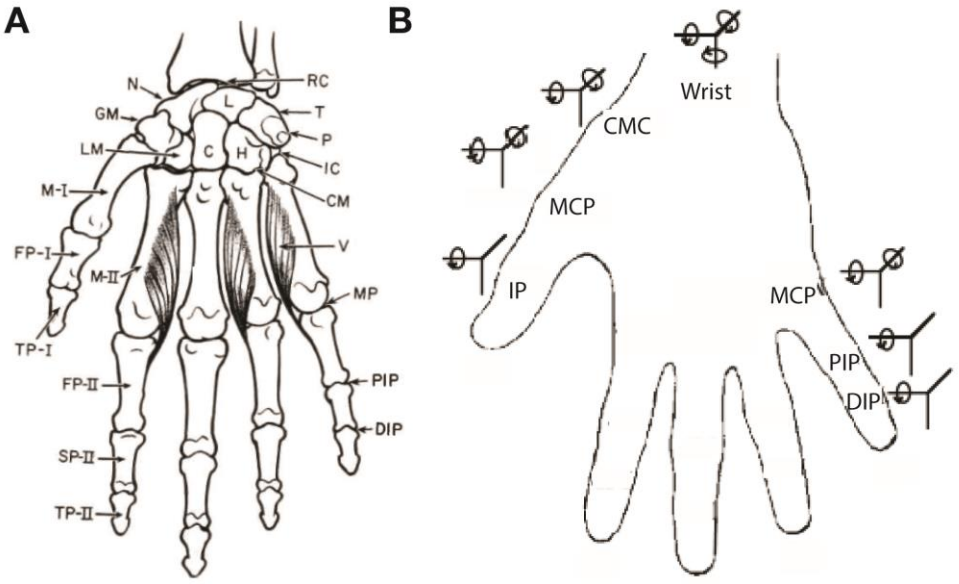

Figure 10: Hand anatomy and degrees of freedom. A. The bone names abbreviated in this figure are the following. Among carpal bones are the greater multangular (GM) also called trapezium, navicular (N), lunate (L), triquetrum (T), pisiform (P), lesser multangular (L), capitate (C), hamate (H). The metacarpal bones are present one in each digit and labelled M-I to M-V. FP, SP and TP correspond to first, second and third phalanges, also called proximal, middle and distal phalanges, respectively. In addition, labelled hand joints are radiocarpal (RC), intercarpal (IC), carpometacarpal (CM or CMC), metacarpophalangeal (MP or MCP), proximal interphalangeal (PIP), and distal interphalangeal (DIP). Figure adapted from Schwarz and Taylor (1955). **B.** DOFs of the hand including 3 at the wrist, and 4 for the joint rotations of each of the five digits, and an additional 5th DOF of the thumb. Translation of the hand in the x , y, and z directions (not shown in figure) adds 3 more DOFs to hand and arm movements. Figure adapted from Holden et al., (1999)

Numerous industrial applications such as patient rehabilitation, gesture based device automation, robot control, sign language detection, virtual reality also require reliable hand tracking. Given these applications, solutions to hand movement tracking have been pursued by the industry and academia, with incremental success over the last five decades. The first glove prototypes including Sayre glove and the MIT-LED glove were developed in the late 1970s and early 1980s, respectively, for hand movement tracking in lab environments (Sturman and Zeltzer, 1994; Dipietro, Sabatini and Dario, 2008). The Sayre glove is a sensor-based glove that used flexible tubes mounted on each digit with a light source and a photocell at opposite ends. The amount of light received by the photocell was converted into the degree of digit flexion. In contrast, MIT-LED glove used cameras to track LEDs mounted on the glove. A camera placed in front of the hand captured the illumination pattern of the LEDs, which was then interpreted as a specific hand gestures. These tracking options were limited to tracking simple hand gestures and were not yet capable of capturing finer digit movements of the hand. Over the years, hand movement tracking techniques using instrumented gloves, as well as camera-based systems have advanced technologically to capture more detailed hand kinematics.

1.5.1 Instrumented data gloves

Different types of instrumented gloves have been developed using optical, stretch, and piezoresistive sensors (Dipietro, Sabatini and Dario, 2008). DataGlove (VPL Research Inc., Redwood City, California) was the first commercially available hand tracking device (Zimmerman and Lanier, 1991). The DataGlove used optical flex sensors (Zimmerman, 1985) mounted on each digit, consisting of a tube that surrounds the digit such that when the digit is extended maximum amount of light is received by the photosensitive detector, and the amount of light received is attenuated with digit flexion. A major drawback of this design is that it does not measure the extent of abduction of digits as has already been pointed out (MacKenzie, 1995). This problem was addressed by the later designs of optical gloves such as the 5DT glove (Fifth Dimension Technologies Inc.,) and the ShapeHand (Measurand

Inc.,). Alternatively, bend sensors whose resistance changes with digit flexion have also been designed (Yun *et al.*, 1997; Simone *et al.*, 2007; Gentner and Classen, 2009) and employed in Cyberglove II (CyberGlove Systems LCC). One variant of Cyberglove II (Figure 11) is a 22-sensor model that uses multiple proprietary flexion and abduction sensors to measure joint angles. Cyberglove II also has a wireless interfacing module which can be conveniently wrapped to the arm to accurately track the human hand (Dipietro, Sabatini and Dario, 2008).

For the investigation of neuronal control of hand movements, the hand tracking solution must also be applicable to non-human primates since many insights on the neuronal underpinnings of grasping movements have been acquired from this animal model (Schieber and Hibbard, 1993; Lemon, 2008; Vargas-Irwin *et al.*, 2010; Dann *et al.*, 2016; Schaffelhofer and Scherberger, 2016; Michaels and Scherberger, 2018).

Figure 11: Illustration of a wireless data glove. Cyberglove II, a commercially available instrumented glove that tracks 22 joint-angles. (http://www.cyberglovesystems.com/cyberglove-ii#specs, accessed Sept 1, 2020)

The first glove designed for non-human primates used 9 flex sensors that were stitched into the pockets of the elastic glove material (Overduin *et al.*, 2010). However these sensors in addition to requiring a prolonged calibration procedure (Overduin *et al.*, 2010) have been found to exhibit a decay in measurement over time (Simone and Kamper, 2005) making them rather unreliable to measure fine changes in bend angle. To address these issues, a kinematic data glove based on electro-magnetic sensors was developed to record hand kinematics of non-human primates (Schaffelhofer and Scherberger, 2012). This solution required only 7 electromagnetic sensors to track 27 DOFs of the macaque hand and the arm. The usage of fewer sensors makes the glove

less bulky and therefore more suitable for use on the macaque hands. Although recording grasp kinematics over a large repertoire of objects and grasps has been demonstrated using the glove (Schaffelhofer and Scherberger, 2016), it requires intense training to get the animals to tolerate it. Furthermore, tracking is limited to the range of the electromagnetic field generator (around 0.5m x 0.5m x 0.5m in the above-cited papers) and ferromagnetic substances cannot be used in the experimental setup. Importantly, appropriate distancing (~30 cm in the above-cited papers) between the field generator and the workspace of the experiment must be maintained to ensure no interference of the generated field with neuronal recordings. Overall, instrumented gloves limit the range of natural hand movements due to the sensors mounted on the digits and impose many restrictions on the experimental setup making it necessary to explore alternative solutions.

1.5.2 Camera based hand tracking systems

Camera based systems have the potential for markerless and even completely contactless hand tracking. With current technology, camera based tracking systems have the potential to record from larger workspace volumes than instrumented gloves. Furthermore, cameras that record at the same or even higher spatial as well as temporal resolution than instrumented gloves are available.

Early camera-based tracking systems used fingertip markers (Davis and Shah, 1994), and colored cloth gloves (Iwai *et al.*, 1996) for rather simple gesture recognition. Although these systems did not require the user to wear instrumented gloves, they still required targeted visual markers or cloth gloves without sensors. The idea of completely markerless and gloveless hand tracking was implemented in Digiteyes (Rehg and Kanade, 1994a). Digiteyes employs a hand model consisting of 3 digits and palm, which was used to estimate the state or the pose of the hand from images acquired from a single camera. Pose estimation in model-based approaches involves searching across candidate configurations of the model to select a configuration which has features that best match the features extracted from the input image (Erol *et al.*, 2007). The three digits hand model of Digiteyes was intended

for use in simple human computer interaction such as gesture based tracking of the computer mouse. An alternative hand model with 27 DOFs that estimated more complex poses of the hand was also integrated with Digiteyes (Rehg and Kanade, 1994b). Features extracted from images acquired simultaneously from 2 cameras were matched with the features of this model for pose estimation. This solution was limited by technology at the time, for example the acquisition frame rate of cameras was around 10 Hz. Furthermore, the models required knowledge of the geometry of the hand to be tracked and worked only under controlled conditions on small subsets of hand poses. Changes in background, increased speed of hand movements, and occlusions among other real world conditions had very strong adverse effects on hand pose estimation. Due to these difficulties in markerless tracking, commercial systems such as the Vicon motion tracking system (Vicon Motion Systems, Oxford, UK) chose passive reflective markers, and Cineplex Behavioral Research System (Plexon Inc., Dallas, Texas) chose tracking colored regions defined by the user. These tracking systems provide the position of each tracked marker in separable data streams making the implementation of post-processing routines easier. Depth sensors such as Kinect (Microsoft Corp., Redmond, Washington), RealSense (Intel Corp., Santa Clara, California), and Leap Motion (Leap Motion, Inc., San Francisco, California) that record depth spectrum along with 2D image of the scene revived research interest in markerless tracking. These depth sensors have been employed successfully for markerless 3D hand (> 20 DOFs) tracking in multiple studies (Sharp *et al.*, 2015; Sridhar *et al.*, 2015) albeit without including any object interactions.

Object interactions are an important part of studies investigating cortical control of hand movements (Baumann, Fluet and Scherberger, 2009; Rouse and Schieber, 2015; Schaffelhofer and Scherberger, 2016). Including object interactions increases the complexity of tracking hand movements for vision-based systems because certain parts of the hand get occluded by the object. Vicon system has been used in experiments that correlated hand kinematics with neuronal data (Vargas-Irwin *et al.*, 2010; Aggarwal *et al.*, 2013). In these two studies a large number of optical markers (n = 29, n = 30 markers were used in the study by Irwin et al., (2010)

and Aggarwal et al., (2013), respectively) and cameras (n = 12, n =18) were required to track DOFs (n = 25, n = 21) in the hand and the arm of non-human primates. Although submillimeter precise tracking was obtained in these studies, the requirement of unobstructed line of sight restricted the studied hand movements making them more unnatural, such as fetching objects suspended on a swinging cable in order to minimize occlusions (Vargas-Irwin *et al.*, 2010).

Vision-based markerless hand tracking including object interactions and occlusion management are increasingly being addressed by solutions employing artificial neural networks (Cao *et al.*, 2017; Simon *et al.*, 2017). These solutions are a promising alternative to track unrestricted and natural hand kinematics.

1.5.3 Deep learning for computer vision

Deep learning allows computational models such as artificial neural networks that are composed of many processing layers to learn different data representations for feature extraction, pattern recognition, and classification purposes (Deng and Yu, 2013; LeCun, Bengio and Hinton, 2015). Artificial neural networks consist of many simple, connected processing units called neurons that produce sequences of real-valued activations (Schmidhuber, 2015). Training corresponds to learning the weights or connection strength between all the connected neurons in the network. In supervised learning, which is the most common form of machine learning (LeCun, Bengio and Hinton, 2015), networks learn the mapping between inputs and outputs from the training examples provided to them. In the remainder of this section, some concepts of deep learning that are relevant for this thesis are introduced along with example deep neural networks used in computer vision applications.

Generally, training neural networks begins with forward propagation, which is carried out as follows. First, inputs belonging to the training dataset are passed through the network. Second, the network generates outputs based on the initial set of weights, which are typically randomly initialized. Third, the network generated outputs are compared to the desired outputs provided in the training set and a loss function (e.g., mean squared error) is employed to estimate the loss or error values.

Following forward propagation, loss is minimized using gradient descent based optimization. The derivatives (or gradients) that are essential for gradient descent in multilayer networks are calculated by implementing backpropagation (Rumelhart, Hinton and Williams, 1986). Weights of the network are updated from the calculated gradients such that the average output error computed on training examples is reduced and the loss value gets closer to a minimum. The forward and the backward propagation steps are repeated for many iterations until the loss reduces to a level that is acceptable for the application. In a variant of gradient descent called stochastic gradient descent, instead of employing the entire training set during every iteration only a small randomly selected subset called a batch is used. Stochastic gradient descent has been demonstrated to yield a good set of network weights surprisingly quickly in contrast to more complicated optimizing procedures (LeCun *et al.*, 2012). Following training, network performance is quantified on a separate set of examples called a test set, which evaluates the generalization capabilities of the network.

Different types of layers are exploited in the design of artificial neural networks. Fully connected layers connect all the neurons of one layer to the next. In contrast, convolutional layers transform activity of one layer to the next by applying two dimensional convolutional or filtering operations. Contrary to fully connected layers, the connections in convolutional layers are more local, and shared. Convolutional neural network (ConvNet) is a category of neural networks that includes convolutional as well as fully connected layers in its architecture and has achieved immense practical success in the field of computer vision in recent years. A ConvNet called AlexNet (Krizhevsky, Sutskever and Hinton, 2012), was the first neural network that outperformed other machine learning techniques in image classification at the 'ImageNet Large Scale Visual Recognition Challenge' (Russakovsky *et al.*, 2015) conducted in the year 2012. The first few stages of AlexNet are composed of convolutional and pooling layers. The task of the convolutional layers is to extract different features present in the data for further processing. Typically a pooling layer follows the convolutional layer to downsize the data by replacing the activations of local groups of neurons with the maximum value of the group. Pooling

operation not only downsizes the data but also provides invariance to small shifts and distortions in data. In the last stages of AlexNet, fully connected layers are used. This design of AlexNet reflects the typical architectural principle of ConvNets, which allows features extracted from input images using different filters to be processed separately in the initial stages and combined later for classification.

Assuming that adding more layers increases the complexity of the features learnt by the network, deeper networks were expected to outperform shallower networks. Contrary to this line of reasoning, He *et al.*, (2016) demonstrated that increasing the depth beyond a maximum threshold resulted in degradation in classification accuracy in deep neural network models. Overfitting was examined as a plausible cause for this degradation of network performance. However, this possibility was ruled out as increased error in deeper networks was present not only on testing dataset but also on training dataset. The problem in training deeper networks is sometimes linked to the magnitude of gradients. Vanishing gradients (Hochreiter, 1991; Bengio, Simard and Frasconi, 1994) for example result in small or no update in weights of neurons in the early layers as the gradients get reduced due to repeated multiplication during back propagation. The identity or skip connections introduced in residual neural networks (ResNets) provided an additional direct path between some of the layers. This solved the problem of training deep architectures making deep neural networks highly scalable (He *et al.*, 2016). ResNet with 152 layers achieved an error rate of 3.57% on the test data of ImageNet database in object classification, which was lesser than that observed in humans (Alom *et al.*, 2018). However, it is important to note that the error rate used for this comparison corresponds to top-5 error rate, which is the percentage of images where the correct output was not among the top-5 most likely outputs. Nevertheless, with its high accuracy in object recognition, ResNet-152 was the top-performer of the 'ImageNet Large Scale Visual Recognition Challenge' in the year 2015. Since then ResNet has been successfully employed in many computer vision related applications (Insafutdinov *et al.*, 2016; Mathis *et al.*, 2018).

1.5.4 ConvNets in hand tracking applications

Several solutions have been proposed for vision based hand tracking by exploiting the advances in deep learning. The first application of ConvNet to reconstruct continuous 3D pose of human hands was made on images captured using depth sensors (Tompson *et al.*, 2014). In this system images were segmented and the hand centered segments were given as inputs to a ConvNet with two convolutional layers followed by three fully connected layers. The outputs of the network were heatmaps and the number of heatmaps corresponded to the number of joints tracked in the study. The peak of heat maps indicated the 2D position of the joints of interest. To obtain 3D from 2D positions, a model fitting procedure was then carried out. Another study inspired by the above idea used a similar network architecture (Ge *et al.*, 2016), however, it avoided the model fitting procedure by first projecting the input depth images onto x-y, y-z, and x-z planes and then training 3 different ConvNets to track joints in 2D across all the three planes. Following this, the predictions of the joints obtained from the 3 planes were combined to estimate the 3D joint positions. The authors compared their ConvNet-based solution to other solutions that did not use neural networks on different challenging academic datasets and demonstrated the better hand tracking capability of their ConvNet. While the above systems employed ConvNets for hand pose estimation on images captured using depth sensors, a much more computationally efficient solution for gesture recognition was proposed by the research group at Google Inc. Their solution is called MediaPipe Hands and it predicts 21 key points of the hand from RGB input images in real-time with high precision even on consumer mobile devices (Zhang *et al.*, 2020). However, the above gesture recognition systems, did not include hand-object interactions as they are intended for applications such as decoding of sign language and other device automation procedures.

To track hand kinematics with higher DOFs and object interactions precisely, it is essential to address the issue of occlusions. A recently developed solution has been demonstrated to estimate hand pose even in the presence of heavy occlusions (Simon *et al.*, 2017). Simon and others trained a ConvNet called convolutional pose

machine (Wei *et al.*, 2016) for the hand tracking task. They recorded from a large number of cameras such that the redundancies in the videos acquired from the cameras can be exploited to automatically generate large training datasets. The main drawback of this solution, however, is that it employs a large number (n = 31) of HD cameras operating in a specially designed panoptic studio (Joo *et al.*, 2019), which is not easy to replicate across lab experiments.

Alternatively, DeepLabCut (DLC) a markerless tracking solution that can be easily applied across different lab experiments was proposed (Mathis *et al.*, 2018).

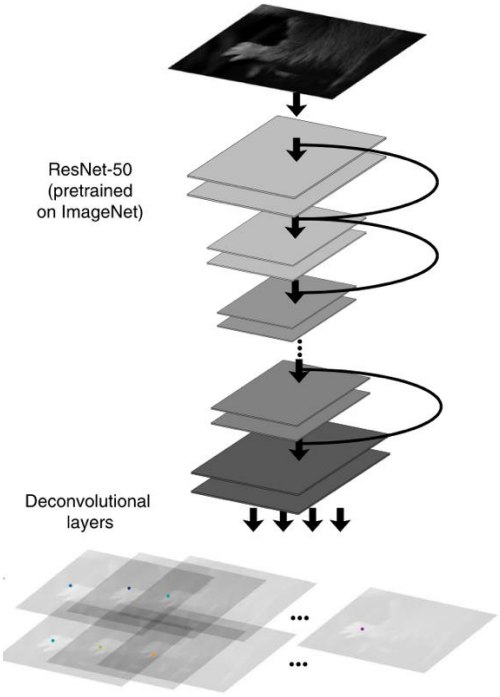

Figure 12: 2D keypoint inference using DLC. Illustration of a ResNet-50 pretrained on ImageNet database applied to infer keypoints of a mouse hand. The network outputs heatmaps which have the same size as the input images. The number of output heatmaps matches the number of keypoints to be tracked and the intensity values of the heatmap correspond to the likelihood values of the associated keypoint. Figure from Mathis *et al.*, (2018)

Using DLC, ResNet pretrained on ImageNet database (Deng *et al.*, 2009) for object recognition task can be retrained to infer keypoints on the hand (Figure 12). This approach of initiating network training from a pre-trained network instead of a network with random weights is referred to as transfer learning. Networks trained for commercial applications on massive datasets can be retrained with relatively small datasets for other tasks sharing common features using transfer learning. DLC

exploits transfer learning to make training possible even with very limited training data (a few hundred examples). It has already been successfully applied to track motor behaviors in several species including mice, flies, horses (Mathis *et al.*, 2018) and also, more recently, in monkeys (Berger, Agha and Gail, 2020). However, the tracked kinematics had lesser degrees of freedom than the primate hand, object interactions were non-existent or very simple, and occlusions were rather minimal. As a part of this thesis, a solution for markerless tracking of 27 DOFs of the primate hand during object interactions was developed that employed only 5 cameras. Importantly, by exploiting constrains of multiview geometry we programmed a procedure to automatically increase the number of keypoint annotations that can be employed for network training. Furthermore, the solution is applicable to track unoccluded keypoints as well as keypoints that are occluded in some of the camera views.

This thesis is divided into two parts. The first part (Chapter 2.1) is dedicated to the study of neuron-level dynamics of oscillatory network structure during sensorimotor transformations. To this end, two macaques were trained to perform a delayed grasping task with randomly mixed instructed and free-choice trials in which a handle had to be grasped with one of two possible grip types. Neuronal activity was recorded in parallel from the fronto-parietal grasping network including the ventral premotor cortex (area F5) and the anterior intraparietal area (AIP) with 64 electrodes chronically implanted in each area. Networks of oscillatory synchrony estimated from spike-field phase locking were characterized to examine their role in behavior dependent information processing.

The second part of this thesis including Chapters 2.2 and 2.3 are dedicated to the development of a markerless tracking paradigm for grasping behavior in primates. The system tracks 22 keypoints of the human hand from videos acquired simultaneously from 5 cameras by exploiting deep learning for computer vision as detailed in Chapter 2.2. Occlusions during object interactions and the availability of a relatively small number of human annotated training data were two of the important challenges addressed. The system was validated on a human grasp

tracking experiment that included 10 objects and a large repertoire of grasps.

In Chapter 2.3, 3D reconstruction of keypoints of the hand tracked in 2D across multiple cameras, which is an important step for computer vision based tracking of hand kinematics is detailed. A Matlab (The MathWorks Inc., Natick, Massachusetts) implementation of a semi-automated 3D reconstruction workflow called pose3d[1] that includes camera calibration, undistortion, and triangulation of keypoints tracked using DLC (or any other 2D tracking software) is presented.

[1] https://github.com/SwathiSheshadri/pose3d

2 Original Articles and Manuscripts

This Chapter contains the following research articles and manuscripts

2.1 Distinct beta and low frequency oscillatory neurons participate in a behavior dependent reconfiguration of the fronto-parietal grasping network in macaques

Swathi Sheshadri*, Benjamin Dann*, Hansjoerg Scherberger. *Equal contribution

Manuscript prepared for submission

Author contributions: S.S., B.D., and H.S. designed and planned the project. S.S., and B.D. analyzed the data. S.S., and B.D. wrote the manuscript. All authors revised the manuscript.

2.2 Precise markerless tracking of the hand during object grasping

Swathi Sheshadri*, Timo Hueser*, Benjamin Dann*, Hansjoerg Scherberger.
*Equal contribution

Manuscript under preparation

Author contributions: S.S., T.H., B.D., and H.S. designed and planned the project. T.H., B.D., and S.S performed the experiment and built the experimental setup. S.S., and T.H. analyzed the data. S.S. wrote the manuscript. All authors revised the manuscript.

2.3 3D reconstruction toolbox for behavior tracked with multiple cameras

Swathi Sheshadri, Benjamin Dann, Timo Hueser, Hansjoerg Scherberger

Published in JOSS: DOI: https://joss.theoj.org/papers/10.21105/joss.01849

Author contributions: S.S., B.D., T.H., and H.S. designed and planned the project. S.S., programmed and documented the toolbox. T.H., B.D., and S.S performed the experiment and built the experimental setup. S.S. wrote the manuscript. All authors revised the manuscript.

2.1 Distinct beta and low frequency oscillatory neurons participate in a behavior dependent reconfiguration of the fronto-parietal grasping network in macaques

Swathi Sheshadri[1,2*], Benjamin Dann[1*], Hansjörg Scherberger[1,2†]

[1]German Primate Center, Kellnerweg 4, 37077 Göttingen, Germany
[2]Faculty of Biology and Psychology, University of Göttingen, 37073 Göttingen, Germany

[*]Equal contribution

[†]Corresponding author. Email: hscherberger@dpz.eu

Manuscript prepared for submission

Author contributions: S.S., B.D., and H.S. designed and planned the project. S.S., and B.D. analyzed the data. S.S., and B.D. wrote the manuscript. All authors revised the manuscript.

Abstract

Oscillatory synchrony is proposed to play an important role in flexible sensory-motor transformations. Thereby, it is assumed that changes in the oscillatory network structure at the level of single neurons lead to flexible information processing. Yet, how the oscillatory network structure at the neuron-level changes with different behavior remains elusive. To address this gap, we examined changes in the fronto-parietal oscillatory network structure at the neuron-level, while monkeys performed a flexible sensory-motor grasping task. We found that neurons formed separate subnetworks in the low frequency and beta bands. The beta subnetwork was active during steady states and the low frequency network during active states of the task, suggesting that both frequencies are mutually exclusive at the neuron-level. Furthermore, both frequency subnetworks reconfigured at the neuron-level for different grip and context conditions, which was mostly lost at any scale larger than neurons in the network. Our results, therefore, suggest that the oscillatory network structure at the neuron-level meets the necessary requirements for the coordination of flexible sensory-motor transformations.

Introduction

One of the most important functions of the primate brain is to transform perceived information into the generation of movements to interact with the environment. This process requires the flexible selection of relevant perceptual information, which needs to be flexibly transformed into movement related activity according to the behavioral demands. Rhythmic co-activations of neurons, referred to as oscillatory synchrony, has been proposed as a mechanism for selective and flexible behavior dependent information processing (Roelfsema *et al.*, 1997; Pesaran, Nelson and Andersen, 2008; Buzsáki, 2010; Nácher *et al.*, 2013; von Nicolai *et al.*, 2014; Bastos, Vezoli and Fries, 2015; Fries, 2015; Wong *et al.*, 2016). Thereby it is assumed that changes in oscillatory network structure at the neuron-level lead to flexible information processing.

The fronto-parietal circuit is known to be strongly involved in flexible sensory-motor transformations (Baumann, Fluet and Scherberger, 2009; Fluet, Baumann and Scherberger, 2010; Michaels *et al.*, 2015; Dann *et al.*, 2016; Michaels and Scherberger, 2018), which are accompanied with changes in oscillatory synchrony (Pesaran *et al.*, 2002; Scherberger, Jarvis and Andersen, 2005). At the area-level within and between these areas, several studies have reported strong beta synchrony during steady states such as periods of rest or movement preparation (Scherberger, Jarvis and Andersen, 2005; Pesaran, Nelson and Andersen, 2008) and a global decrease in beta synchrony was shown to be predictive of movement initiation (Canolty, Ganguly and Carmena, 2012; Womelsdorf, Westendorff and Ardid, 2013). Therefore, beta is assumed to be related to the maintenance of status quo (Engel and Fries, 2010) and to be involved in the transition from movement planning to execution. However, the involvement of beta in flexible information processing remains debated because several studies reported only small or even no context or condition dependent changes in beta synchrony (Scherberger, Jarvis and Andersen, 2005; Pesaran, Nelson and Andersen, 2008; Engel and Fries, 2010; Haegens *et al.*, 2011, 2017; Dean, Hagan and Pesaran, 2012; Martínez-Vázquez and Gail, 2018). In contrast to beta synchrony, recent

studies have related low frequency synchrony within and between fronto-parietal areas to movement execution (Churchland *et al.*, 2012; Elsayed *et al.*, 2016; Martínez-Vázquez and Gail, 2018).

In the fronto-parietal network at the neuron-level, a few studies provided evidence that only a distinct class of neurons is oscillatory synchronized in the low frequency and beta bands, while many neurons are not oscillatory synchronized (Dean, Hagan and Pesaran, 2012; Dann *et al.*, 2016; Wong *et al.*, 2016). These oscillator-neurons were strongly synchronized within and between areas forming a subnetwork, which may be central for flexible information processing (Dann *et al.*, 2016).

These previous approaches that either examined temporal dynamics at the area-level or the network structure at the neuron-level of oscillatory synchrony fall short, however, to capture how changes in the oscillatory network structure at the neuron-level are related to flexible transformation processes. In this regard, several fundamental questions remain elusive. During behavior dependent processes, it is unknown whether the same population of neurons synchronizes in different oscillatory frequencies or whether distinct subpopulations of neurons exist that only synchronize in one particular frequency band (Figure 1A). Furthermore, it is also unknown whether at the level of neurons in the network, different frequencies are associated with different behavioral states, and whether different frequencies in the network are therefore mutually exclusive (Figure 1B). Finally, it is unclear whether fine-scale reconfigurations of the oscillatory network structure are present at the neuron-level for different conditions (e.g., different sensory inputs, movement plans or movement executions), which are lost at larger scales such as the area-level (Figure 1C). This is of particular importance because condition dependent changes of the oscillatory network structure are a prerequisite for flexible information processing by oscillatory synchrony.

Here, we investigated changes in the oscillatory network structure at the neuron-level over time and across different conditions during flexible sensory-motor transformations. For this purpose, we simultaneously recorded large populations of

single neurons and LFPs of the fronto-parietal grasping network (F5 and AIP).

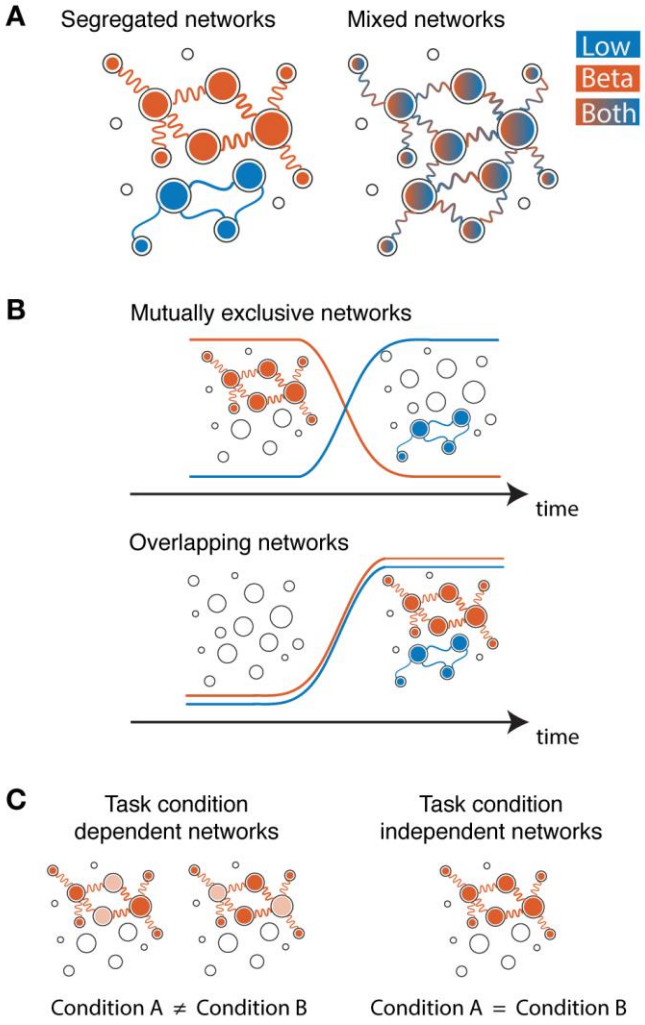

Figure 1: Hypothesized oscillatory network structure dynamics of the fronto-parietal grasping circuit. (A) Illustration of the hypothesis in which the groups of neurons participating in the two behavior-relevant frequency bands (low and beta) were examined to understand if the same or different neurons participate in oscillatory synchrony in the two frequency bands. Distinct groups of neurons participating in the two frequency bands result in segregated networks, while overlapping groups result in mixed networks. **(B)** In this hypothesis, we focused on the temporal dynamics of the networks. We examined if the networks in the low frequency and beta bands were active at different times mutually exclusively or at the same time overlappingly. **(C)** In this hypothesis, we examined the behavior dependence of networks underlying task conditions. Different networks underlying different task conditions result in task condition dependent networks, while the same networks underlying different task conditions result in task condition independent networks.

In agreement with previous studies, we found strong 3-6 Hz low frequency and 17-35 Hz beta synchrony across the network. However, individual neurons were predominantly either beta or low frequency synchronized with large parts of the network forming separate frequency specific subnetworks, with beta neurons

predominantly located in AIP and the low frequency neurons in F5. Over the time course of the task, beta neurons were strongly synchronized with the network during steady states such as the delay epoch and weakly synchronized during active states such as movement execution, while low frequency neurons showed the opposite temporal profile. This suggests that both frequencies are mutually exclusive in the network at the neuron-level. Furthermore, individual neurons showed strong condition dependent changes in oscillatory synchrony with the network in both frequencies, despite only small detectable changes at the area-level. Closer examinations of the entire oscillatory network structure revealed that more than 80% of condition dependent changes can only be captured at the fine-scale network-level and that changes of individual connections even accounted for more than 25% of these changes. Our results thus suggest that fine-scale reorganizations of frequency specific subnetworks of neurons are the core of behavior dependent information processing by oscillatory synchrony.

Results

Behavioral tasks, single neurons and LFPs recordings

To study changes in oscillatory synchrony during transformation processes, we trained two monkeys (S and Z) to perform a delayed grasping task. In this task, the monkeys were either instructed to grasp a target with one of the two possible grip types (power and precision), or could choose freely between the grips, as described in detail in previous studies (Michaels *et al.*, 2015; Dann *et al.*, 2016) (Figure 2A). In the instructed-context monkeys were visually cued by one of two discs displayed on a monitor to perform the associated grip type. In the free-choice context both discs were displayed and monkeys could choose freely between the grip types. Both monkeys learned to perform the task with high accuracy and high trial counts (percentage of successful trials: 95 ± 0.01 SD % and 96 ± 0.03 SD %, number of successful trials: 730 ± 106 SD and 722 ± 167 SD for S and Z, respectively).

During the task, we simultaneously recorded large populations of single neurons and LFPs from the ventral premotor cortex (area F5) and the anterior intraparietal area (AIP), which are both part of the fronto-parietal grasping network.

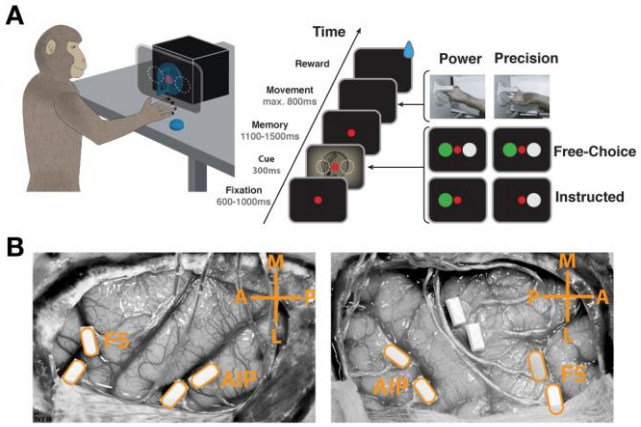

Figure 2: Experimental design and location of implanted floating micro-electrode arrays. (A) In this delayed grasping task paradigm monkeys were cued to grasp a handle with one of two different grip types displayed on a monitor and superimposed onto the handle. To start the trial the monkeys had to fixate a red disk for 600–1000ms (Fixation). Then, the cue epoch lasting 300ms followed, when either ('Power') a green disk was presented on the left indicating a power grip, ('Precision') a grey disk was presented on the right indicating precision grip, or ('Free-choice') both disks were presented indicating a free-choice between power and precision grips. After the cue was turned off, the monkeys were required to remain steady during the memory epoch (duration: 1100–1500ms). Then, the fixation dot was turned off (go-signal), indicating to the monkey to execute the grasp movement (maximum duration: 800ms). **(B)** Pictures of implanted floating micro-electrode arrays in monkey Z (left) and monkey S (right). Animals were implanted with 4-6 floating Microprobe arrays, in areas AIP and F5 (M1 data from monkey S not used for this study).

Recordings were made from two chronically implanted 32-channel microelectrode arrays per area (64 channels per area and 128 channels in total; Figure 2B; see STAR Methods). For all the following analyses, 6 recording sessions from monkey S and 3 from monkey Z were used. To ensure a stable estimate of oscillatory synchrony of individual neurons over time and across conditions, only well-isolated neurons (see STAR Methods) with at least 70 spikes during all 800ms time windows of all conditions of the task were used for all further analyses. This resulted in an average number of neurons per recording session of 15.6 ± 2.6 and 14.3 ± 3.5 for area F5 and 24.1 ± 2.8 and 8.6 ± 0.5 for area AIP for S and Z, respectively. Noisy LFP-sites were excluded from all analyses and to ensure that the LFP signals are local, we removed the common recording reference by using linear regression (see STAR Methods). This

resulted in an average number of LFP-sites per recording session of 54.8 ± 2.7 and 52.0 ± 1.73 for area F5 and 56.6 ± 1.2 and 58 ± 1.0 for area AIP for S and Z, respectively.

Area-level spike-field phase locking over time and across conditions

We characterized the behavior dependent changes in oscillatory synchrony in the fronto-parietal network by estimating the strength of spike-LFP phase locking between all neuron-LFP pairs per recording session employing the method PPC (see STAR methods) (Vinck *et al.*, 2010, 2012). PPC was computed over the time course of the task with a sliding window of 800ms and separate per condition (step size of 100ms, 23 time windows x 4 conditions). Using PPC to estimate phase locking has the advantage that PPC is unbiased by spiking rate. Thus, spike-LFP phase locking estimated by PPC allows the comparison of different neurons, LFP-sites, time points and conditions without the confounding influences of rate changes.

Figure 3A shows all PPC spectra of an example AIP neuron with all simultaneous recorded LFP-sites during memory epoch for all four task conditions of a representative recording session from monkey S. The example neuron was strongly phase locked in the beta band (18 - 35 Hz) with most of the LFP-sites in AIP and some of the LFP-sites in F5. Across all neuron-LFP pairs of both areas, time windows, conditions and recording sessions, neurons were strongly phase locked in the low frequency (3 - 6 Hz) and beta bands for both monkeys (Figure 3B; see Figure S1 for an example F5 neuron phase locked in the low frequency band with many of the simultaneously recorded LFP-sites). In a direct comparison of both areas, F5 neurons were on average predominantly phase locked in the low frequency band and AIP neurons predominantly phase locked in the beta band with LFP-sites from both areas for both monkeys. LFP-site phase locking per area averaged across all neurons showed qualitatively the same but weaker area bias in the beta band, and was similar between areas in the low frequency band (Figure S1B). These results therefore suggest that phase locking of neurons is more area specific than phase locking of LFP-

sites.

To validate whether the observed low frequency and beta band phase locking was significant and to test whether significant phase locking was present in other frequency bands, we used cluster-based surrogate statistics (see STAR methods). For this purpose, first, surrogate PPCs were generated for every neuron-LFP pair by randomly permuting LFP trials within each behavioral condition, while preserving the total number of spikes and LFP power per frequency for every condition (see STAR methods). Second, each PPC spectrum per neuron-LFP was tested against the corresponding surrogate PPC spectra employing cluster-based statistics. As expected by the average phase locked strength per area (Figure 3B), we found a higher number of significant neuron-LFP phase locking pairs in the low frequency band with neurons predominantly from F5 and in the beta band with neurons and LFP-sites predominantly from AIP (Figures 3C and S1C).

Given strong low frequency and beta band phase locking in the fronto-parietal grasping network, we next examined whether and how phase locking at the area-level changes over the time course of the task. Figure 3D depicts the average PPC spectrograms across all neuron-LFP-pairs of all neurons per area, conditions, and recording sessions (see Figure S1D for LFP-sites per area). In both areas beta phase locking of neurons and LFP-sites was strong during fixation and memory epochs, weak during cue epoch, and nearly absent during movement epoch. Conversely, low frequency phase locking was strong during cue epoch, strongest during movement epoch and nearly absent during fixation and memory epochs. In agreement with the time averaged phase locking results above, AIP neurons and LFP-sites were stronger beta phase locked and F5 neurons stronger low frequency phase locked over the time course of the task.

We further determined the extent to which beta and low frequency phase locking at the area-level differed between the four conditions of the task. Displayed in Figure 3E are the average PPC spectra per condition across all neuron-LFP-pairs of all neurons per area, time windows, and recording sessions (see Figure S1E for LFP-sites per area). Except for small condition dependent differences in low

frequency phase locking of F5 neurons from monkey S and AIP neurons from monkey Z, the average phase locking of neurons and LFP-sites was highly similar across conditions in both frequency bands and areas.

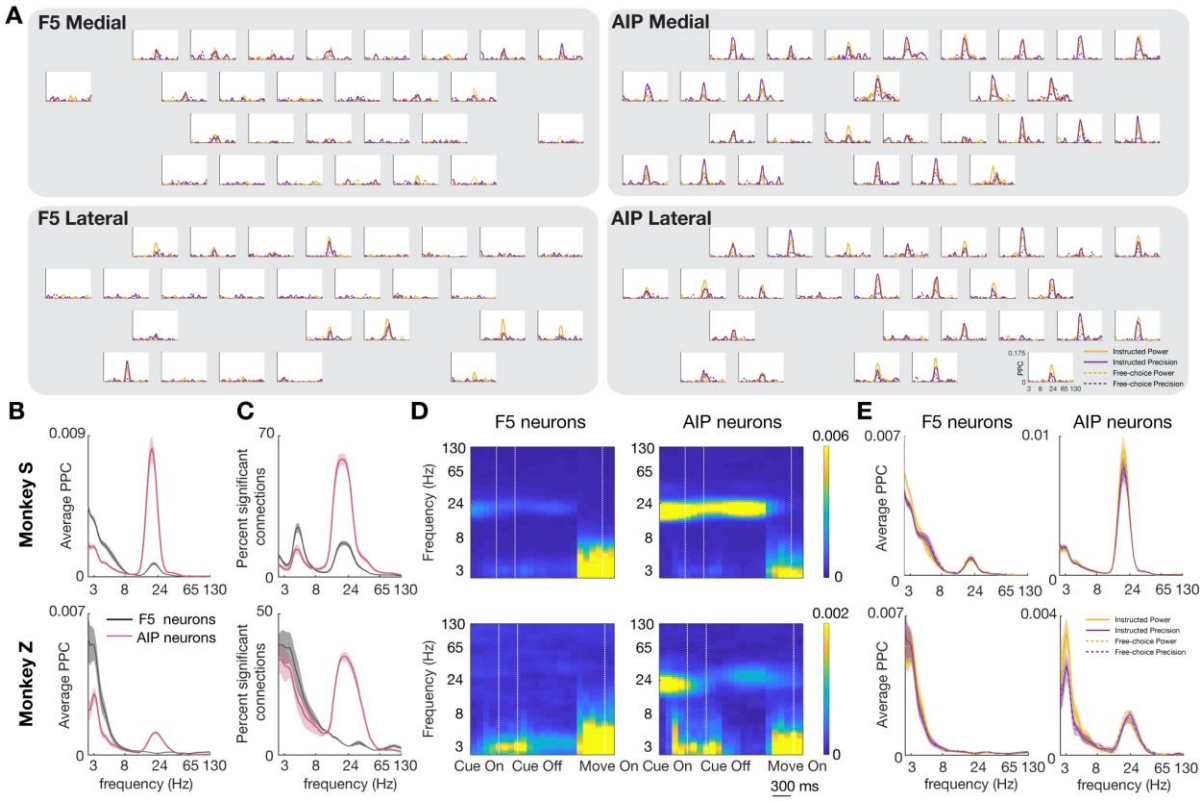

Figure 3: Example spectra and the spatial, temporal and task condition specific differences in neural population averages. (A) PPC spectra during memory epoch of a representative neuron recorded from area AIP with spatially distributed LFP signals recorded from F5 and AIP in monkey S. **(B)** Population average phase locking in PPC spectra of neurons recorded from areas F5 and AIP with all LFP signals. Line shadings indicate standard error across recording sessions. **(C)** Percentage of significant phase locking (identified using cluster-based surrogate tests, see STAR methods) of neurons recorded from F5 and AIP with all LFP signals. Line shadings show standard error across recording sessions. **(D)** PPC spectrograms of neurons recorded from F5 and AIP with LFP signals from both areas to illustrate time-resolved phase locking (100ms resolution). In the PPC spectrograms, data were clipped at 95th percentile for visualization purposes. **(E)** PPC spectrograms of neurons recorded from F5 and AIP with LFP signals from both areas illustrating condition resolved phase locking. (in (B), (C), (D) and (E) results are shown for monkeys S and Z separately).

Taken together, we found strong low frequency and beta spike-LFP phase locking in the fronto-parietal grasping network, in agreement with previous studies (Scherberger, Jarvis and Andersen, 2005; Pesaran, Nelson and Andersen, 2008; Haegens *et al.*, 2011; Churchland *et al.*, 2012; Martínez-Vázquez and Gail, 2018). On average, AIP neurons were predominantly beta phase locked and F5 neurons

predominantly low frequency phase locked. Yet, a small but significant number of F5 and AIP neurons were also beta and low frequency phase locked, respectively.By contrast, phase locking of the LFP-sites in both frequency bands was less different between the areas. One explanation for the higher degree of frequency specificity of neurons than LFP-sites could be that different populations of neurons are phase locked in the beta and low frequency bands with LFP-sites of both areas (Figure 1A). In this case, neurons would form frequency specific subnetworks, which consist of mostly neurons from one area, but also some neurons from the other area. Thus, phase locking of LFP-sites would be less frequency specific because they reflect the average phase locking of all surrounding neurons.

Over the time course of the task at the area-level beta phase locking was strong during fixation and memory epoch, while low frequency phase locking was strong during the cue und movement epoch. This raises the question whether beta phase locking is related to steady states and low frequency phase locking to active observance and movement and therefore phase locking in the two frequency bands is mutually exclusive during sensory-motor transformations (Figure 1B). If phase locking is mutually exclusive in the two frequency bands, both frequencies should be clearly separated in the network of neurons over the time course of the task.

Average beta and low frequency phase locking at the area-level was highly similar for the different grip as well as context conditions of the task. This finding is in accordance with previous studies also showing little to no condition dependent differences in oscillatory synchrony (Pesaran, Nelson and Andersen, 2008; Haegens *et al.*, 2017). Suppose that the phase locking in one frequency of individual neurons changes independently for different conditions, this would result in fine-scale reconfigurations of the phase locking network that are lost at any scale larger than neurons. Therefore the possibility remains that condition dependent reconfigurations in the beta and low frequency phase locking network exist at the neuron-level (Figure 1C).

However, the investigation of these three fundamental questions (Figure 1) requires network analyses at the neuron-level. First, because simultaneous

recordings of neurons with the same LFP-sites are necessary in order to compare their phase locking with each other. Second, because simultaneous recordings of many LFP-sites are necessary to ensure that neuron-level phase locking in both behavior related frequencies is detectable in the same network.

Separate populations of low frequency and beta phase locked neurons

The large number of simultaneously recorded neurons and LFP-sites allows to examine the relationship between low frequency and beta phase locking at the level of neurons in the network. Based on this, we first examined the question whether the same or different populations of neurons are phase locked in the two frequency bands (Figure 1A).

Displayed in Figure 4A is the time and condition averaged anatomical networks of spike-field phase locking separately for the 3 – 6 Hz low frequency and 18 – 35 Hz beta bands of one representative recording session from monkey S (see Figure S1A for one representative recording session from monkey Z). Interestingly, the strength of phase locking per neuron averaged over all LFP-sites was highly heterogeneous across the population of neurons in both frequency bands. For a better comparison of low frequency and beta phase locking, we depict both networks as connectivity matrices with the average strength of phase locking per neuron and per LFP-site at the corresponding edges (Figure 4B, see Figure S2A for a representative recording session from monkey Z). As expected, strong beta phase locked neurons were predominantly located in AIP and strong low frequency phase locked neurons were predominantly located in F5. However, we also observed weakly beta phase locked neurons in F5 and weakly low frequency phase locked neurons in AIP. In direct comparison of the low frequency and beta phase locking networks, neurons strongly phase locked in one frequency seem to be weakly phase locked in the other frequency. LFP-sites, in contrast, were more uniformly phase locked and seem to have a clear overlap in both frequency bands.

To examine whether beta and low frequency neurons belong to the same or

different populations of neurons, we compared the average beta and low frequency phase locking per neuron across all recordings separately for the two monkeys (Figure 4C). Note that the comparison was done separately for both monkeys because the average strength of beta and low frequency phase locking was different between monkeys. For both monkeys, neurons with higher low frequency phase locking were weakly beta phase locked and vice versa. We calculated the cosine similarity between low frequency and beta phase locking of all neurons to quantify their degree of overlap. Cosine similarity between frequencies was low for both monkeys (0.37 and 0.27 for S and Z, respectively). However, a low cosine similarity does not necessarily indicate that low frequency and beta neurons belong to different populations, because the distributions of phase locking of neurons in both frequency bands can influence the results. In particular, the heavy-tailed distribution of phase locking with most neurons weakly phase locked and a few neurons very strongly phase locked, such as in this case, could lead to a large variability of cosine similarity. We therefore used permutation statistics preserving both phase locking distributions to test whether neurons phase locked in the two frequencies belong to different populations (see STAR Methods). Cosine similarity was smaller than expected by chance for both monkeys (2-sided permutation test, $p < 0.0001$ for S and Z), suggesting that low frequency and beta phase locked neurons belong to separate populations.

We next examined whether low frequency and beta phase locking is also separated at the LFP-level. In contrast to neurons, cosine similarity between low frequency and beta phase locked LFP-sites was high (Figure S3A; 0.69 for and 0.72 for S and Z, respectively) and significantly above chance (2-sided permutation test, $p < 0.0001$ for S and Z) suggesting that low frequency and beta phase locking overlaps at the LFP-level.

Together, these results suggest that fronto-parietal low frequency and beta phase locked neurons form separate subnetworks (Figure 1A). Interestingly, the separation of the frequency specific subnetworks could only be detected at the neuron-level and not at the LFP-level. This suggests that both subnetworks contain neurons from the same local populations and are therefore already mixed at the LFP-level.

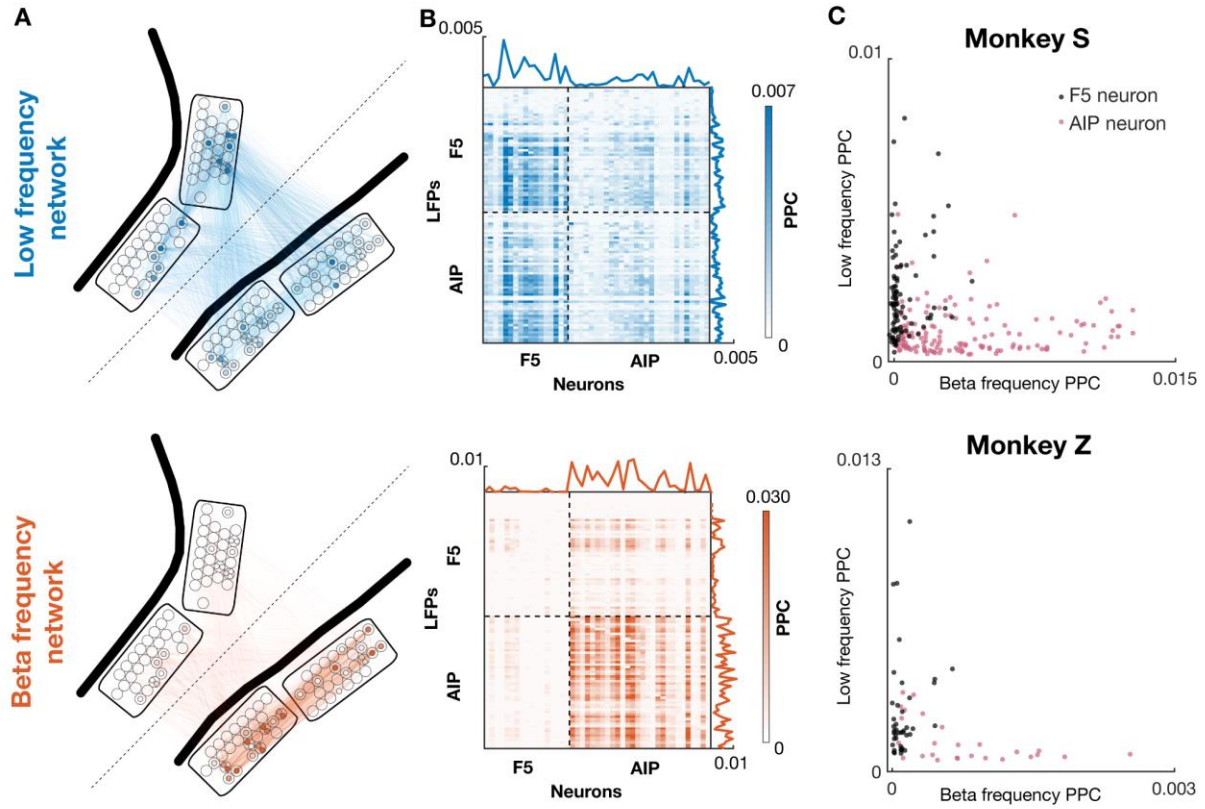

Figure 4: Segregated groups of neurons participate in the low frequency and beta networks. (A) Schematic illustrating average PPCs over all task conditions and time windows in an anatomical representation. In these networks, the four rounded rectangular boxes correspond to the arrays implanted of F5 and AIP, with the dashed line used to separate the two areas. The inner circles correspond to neurons and the outer circles correspond to LFP signals. The thickness of the lines drawn between neurons and LFPs reflects the strength of phase locking between them. The graded color-filling of the neurons shows the average PPC value of the corresponding neuron. **(B)** Networks in (A) illustrated in the format of connectivity matrices with columns representing neurons, rows representing LFP-sites and the graded color scale used to represent the strength of phase locking between neuron-LFP pairs in the network. Average PPC values associated with each neuron and LFP signal are plotted along the top and right margins of the connectivity matrices, respectively. (A) and (B) illustrate networks corresponding to a representative recording session from monkey S and PPC values in the network illustrations were clipped at 99th percentile for better visualization by de-emphasizing outliers. **(C)** Comparison of average low frequency and beta PPC values for neurons from all recording sessions shown separately for monkeys S and Z with color coding to highlight area specificity.

Low frequency and beta phase locking are mutually exclusive over time

We next addressed the question whether beta and low frequency phase locking are associated with different behavioral states at the neuron-level in the network and therefore whether they are mutually exclusive during sensory-motor transformations (Figure 1B). As mentioned above, several studies have related different oscillatory

frequency bands with different behavioral states (Engel and Fries, 2010; Churchland *et al.*, 2012). However, it has not been investigated yet, whether the relationship of different behavioral states with different frequencies also holds true at the level of neurons in the network. For example, if two spatially overlapping populations of neurons are phase locked with the network in the same phase during one behavioral state and with random or opposing phases during the other behavioral state, analyses of brain signals coarser resolved then neurons would detect a decrease in phase locking for the second state. Therefore, analyses at the network-level with neuron-resolution are required to investigate behavioral state dependent differences in oscillatory synchrony.

Displayed in Figures 5A and S2B are the low frequency and beta networks of one representative recording session per monkey over the time course of the task averaged across all conditions per spike-field pair. Across the network, low frequency phase locking was weak during fixation, increased slightly around cue epoch, was weak again during memory epoch and was strongest around movement. Beta phase locking in contrast was strong during fixation, decreased slightly around cue, was strong again during memory epoch, and was nearly absent around movement. Average spike-field phase locking of all spike-LFP pairs over the time course of the task confirmed the observation that strong low frequency phase locking is associated with active observance and movement and strong beta phase locking with steady states during fixation and memory epochs (Figure 5B).

The finding that low frequency and beta phase locking is associated with different behavioral states indicates that phase locking in both frequencies is mutually exclusive in the network at the neuron-level. To examine this assumption, we directly compared beta and low frequency phase locking of all neurons over the time course of the task separately per monkey (Figure 5C). Neuron-level phase locking over time was clearly separated into the two frequency bands and the cosine similarity of beta and low frequency phase locking was significantly below chance level for both monkeys (0.06 and 0.12 for S and Z, respectively; 2-sided permutation test, $p < 0.00001$ and $p = 0.016$ for S and Z, respectively).

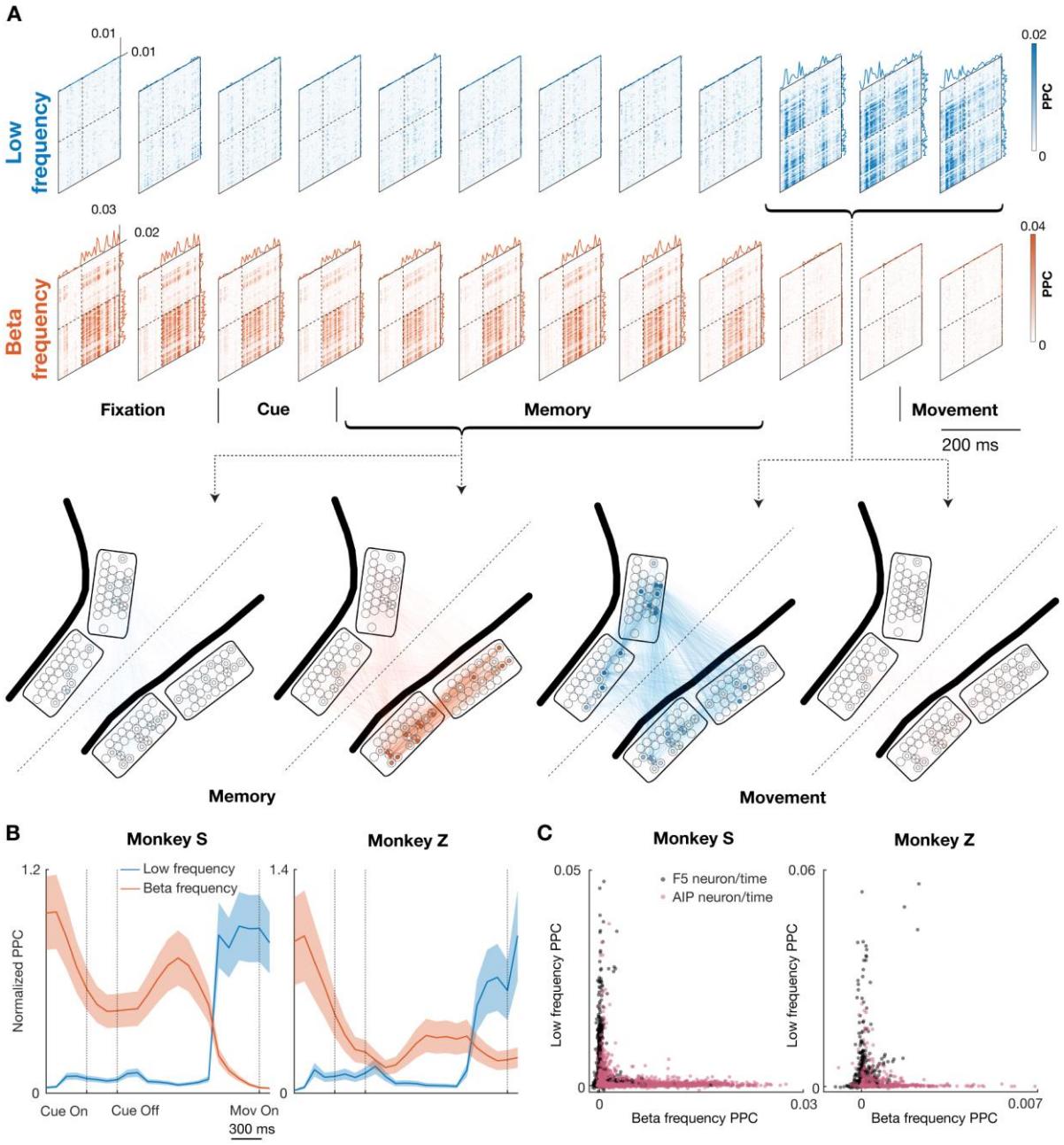

Figure 5: Low frequency and beta networks active strongly during different behavioral epochs. (A) Network representation over time for an exemplar recording session from monkey S shown at a temporal separation of 200ms (networks were computed at a temporal separation of 100ms, only every second network shown here). Low frequency and beta networks from time windows corresponding to peak beta phase locking (800ms time window centered around 900ms after cue onset) representing memory epoch and peak low frequency phase locking (800ms time window centered around 100ms before movement onset) representing movement epoch shown in anatomical network illustration. PPC values in the network illustrations were clipped at 99th percentile for visualization. **(B)** PPC values corresponding to a time window averaged over all neurons, LFPs and task conditions. Normalized low frequency and beta PPC values averaged over recording sessions to compare temporal activation profiles (Line shadings indicate standard error across recording sessions). **(C)** Comparison of average low frequency and beta PPC values corresponding to neurons at different time windows from all recording sessions. Results shown for monkey S and Z separately in (B) and (C).

In comparison to time averaged neuron-level phase locking (Figure 4C), the separation into the two frequency bands was stronger for time resolved neuron-level phase locking implying that phase locking in both frequencies is separated in time. If low frequency and beta phase locking are mutually exclusive over time, this relationship should also be present for LFP-sites. In agreement with this assumption and in contrast to time averaged LFP-site phase locking (Figure S3A), LFP-site phase locking over time was also separated into the two frequencies (Figure S3B; 0.12 and 0.26 for S and Z, respectively; 2-sided permutation test, $p < 0.00001$ for monkeys S and Z)

Together, these results suggest that at the level of neurons in the network low frequency phase locking is associated with active behavioral states and beta phase locking with steady behavioral states, in agreement with previous literature (Engel and Fries, 2010). Furthermore, the clear temporal separation of both frequencies at the neuron- and LFP-level suggest that low frequency and beta phase locking are mutually exclusive in the network (Figure 1B).

Low frequency and beta networks reconfigure for different conditions

Given segregated low frequency and beta phase locked subnetworks of neurons active during different states of sensory-motor transformations, the question remains whether these subnetworks are the same or reconfigure at the neuron-level for different grip and context conditions (Figure 1C). As mentioned earlier, it is possible that despite a similar condition dependent phase locking at the area-level (Figure 3E) the phase locking of individual neurons changes independently for different conditions, which can only be examined at the fine-scale network-level with neuron-resolution.

To address this question, we first examined whether and at which times neuron-level beta and low frequency phase locking was different between task conditions. We quantified condition dependent differences by estimating the neuron-level phase locking variance across conditions and then compared the variance at

each time window with surrogate data variance. Surrogate phase locking networks were generated by permuting trials from all four conditions before estimating PPC, thus preserving the total number of spikes, LFP power, and the amount of low frequency and beta phase locking per neuron-LFP pair (see STAR Methods). The neuron-level phase locking variance was significantly higher between conditions around cue and movement epochs in the low frequency band and during memory epoch in the beta band for both monkeys (Figure 6A; cluster-based surrogate test, $p < 0.007$). However, visual inspection of the time averaged phase locking networks during the three significant periods showed only small differences between conditions (see Figure 6B, and Figure S2B for one representative recording session from monkey S and Z, respectively). This observation is reinforced by a high cosine similarity between all pairs of conditions during all three significant epochs in the corresponding frequency bands (Low frequency cue: 0.71 ± 0.08 and 0.92 ± 0.02, beta frequency memory: 0.93 ± 0.01 and 0.81 ± 0.05, low frequency movement: 0.86 ± 0.06 and 0.80 ± 0.15 for S and Z, respectively). Cosine similarity was significantly above chance in all cases except for grip type differences in low frequency phase locking during movement for monkey Z (2-sided permutation test, $p < 0.05$ Bonferroni corrected for the number of condition pairs).

The supposed discrepancy between significant condition dependent phase locking at the neuron-level and high cosine similarity could be explained by a large amount of condition independent phase locking and a smaller but significant amount of condition dependent phase locking. In agreement with this assumption, condition differences only explained on average 23.5% of phase locking variance at the neuron-level (low frequency cue period: 42 ± 12 SD % and 15 ± 7 SD %, beta memory period: 11 ± 3 SD % and 25 ± 12 SD % of, and low frequency movement period: 21 ± 7 SD % and 27 ± 3 SD % for S and Z, respectively). However, a few neurons showed strong condition dependent changes in phase locking, while most neurons did not (see, for example, the encircled neurons in Figure 6C and S4). To examine condition dependent differences in phase locking across neurons, we computed the difference in phase locking per neuron for grip-type, context and grip-context-interaction during

all three epochs separately per monkey (Figure 6D).

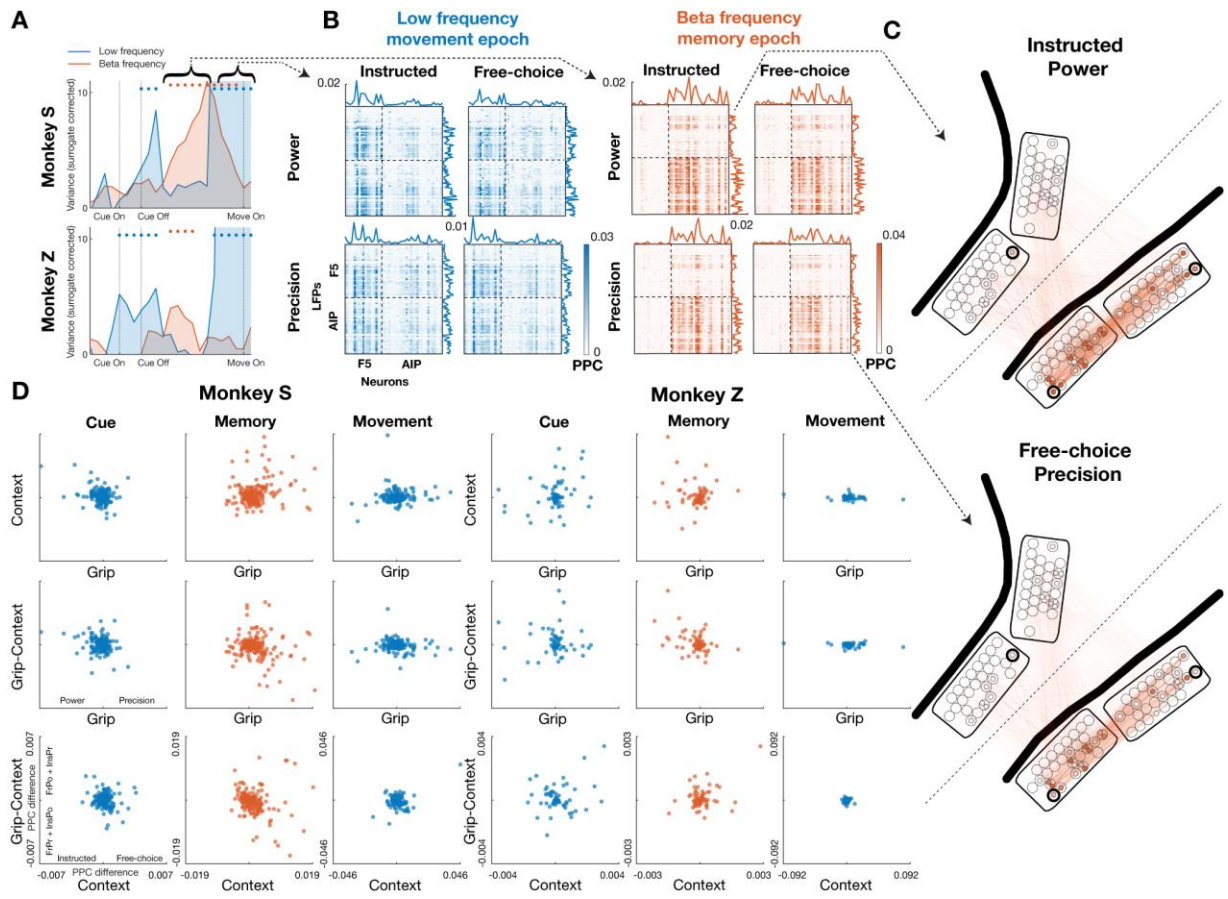

Figure 6: Task condition dependent differences in the low frequency and beta networks. (A) Variance in the neuron-level phase locking corrected by subtracting the average and dividing by the standard deviation of the variances computed over 1000 surrogate networks generated by shuffling condition identity before calculating PPC. Dots indicate time windows when significant condition dependent differences were detected in the low frequency (blue dots) and the beta bands (orange dots). **(B)** Low frequency networks during movement epoch (time windows selected based on the observed significant differences across conditions in (A): 800ms duration windows centered from 300ms before movement onset to 100ms after movement onset) and beta frequency networks during memory epoch (800ms duration windows centered from 600ms after cue onset to 1200ms after cue onset) for the four task conditions. **(C)** Exemplar differences between instructed power and free precision task conditions during memory epoch shown for beta frequency using anatomical network representations. Black circles drawn around select neurons highlight large differences in the average PPC values between the two task conditions. In (B) and (C), PPC values were clipped at 99th percentile for visualization purposes. **(D)** Grip, context and grip-context interaction effects observed at the level of neurons in the network in the low frequency band during the cue and movement epochs and in the beta frequency band during the memory epoch. Grip effect (Power vs. Precision) was determined by averaging PPC values over the two context types in the task for each grip type and subtracting PPC values corresponding to "Precision grip" from PPC values corresponding to "Power grip". Context effect (Instructed vs. Free-choice) was determined by averaging PPC values over the two grip types in the task and subtracting PPC values corresponding to "Free-choice" task from the "Instructed" task condition. Interaction effect was determined by adding PPC values corresponding to "Free-choice Precision and Instructed Power" and subtracting it from the summed PPC values corresponding to "Free-choice Power and Instructed Precision" task conditions.

We found that phase locking changes across neurons were in general heterogeneously

distributed with most neurons showing little to no condition dependent differences, and a few neurons showing strong condition dependent differences. Interestingly, neurons showed comparable differences for grip, context and grip-context interaction during cue and memory period in the low frequency and beta band, respectively, while neurons showed almost exclusively grip-type differences during the movement period in the low frequency band. Behavior dependent changes in phase locking of LFP-sites were more homogeneously distributed in comparison to neurons during the same three periods in the corresponding frequency bands (Figure S3C). Nonetheless, LFP-sites showed a similar ratio of grip, context and grip-context differences in phase locking as neurons during all three periods.

The results above show that grip type and context dependent changes in phase locking are present in the fronto-parietal low frequency and beta subnetworks at the neuron-level. Therefore, these findings confirm that condition dependent reconfiguration of the phase locking network structure at the neuron-level can be present despite little to no condition dependent differences at the area-level (Figure 3E). A possible explanation for this apparent discrepancy could be the comparable amount of condition dependent weaker and stronger phase locked neurons (Figure 6D), which would cancel each other out at any larger scale than neurons. This explanation is supported by the finding that condition dependent differences for LFP-sites were similar but less pronounced compared to neurons.

Interestingly, grip and context related changes in phase locking were present during the cue and memory period, but only grip related changes were present during the movement period. To solve the given task both context and grip information must be processed and transformed differently during the cue and memory period in order to ultimately perform the same two grasp movements regardless of the context. Thus, the found changes in phase locking for grip and context during the three periods accurately reflect the expected behavioral demands.

Neurons with strong phase locking show stronger condition dependent changes

The high level of heterogeneity at the neuron-level of overall phase locking (Figure 4) and condition dependent changes in phase locking (Figure 6) raises the question how both parts of phase locking are related. If overall weakly phase locked neurons show the strongest behavior dependent changes in phase locking, it would suggest that changes in the activation of these neurons indirectly cause phase locking with the core of strongly phase locked neurons. If overall strongly phase locked neurons show the strongest behavior dependent changes in phase locking, it would suggest that the core of strongly phase locked neurons directly causes changes in phase locking.

To investigate this relationship, we first estimated the distribution of overall phase locking per neuron over time and across conditions. Neuron-level phase locking was heavy-tailed distributed in both frequency bands for both monkeys (Figure 7A). Thereby, the majority of neurons was significantly more weakly phase locked and a few neurons significantly more strongly phase locked with the network than expected by chance with the same spatial distance dependent decrease in phase locking (network-level cluster-based permutation test, $p < 0.05$, see STAR Methods).

Given the presence of weakly and strongly phase locked neurons in both frequencies, we next investigated the relationship between these populations and behavior dependent changes in phase locking. Figure 7B depicts phase locking per neuron for all four conditions sorted by overall phase locking during all three significant periods in the corresponding frequency band of one representative dataset from monkey S and Z. In all example cases, neurons with higher overall phase locking showed stronger condition dependent changes in phase locking. To compare the dependency of overall and condition dependent changes in phase locking across recording sessions, we first computed the variance in phase locking between conditions per neuron and second resampled and averaged the variance sorted by the overall phase locking to compensate for the different number of neurons across recording sessions (Figure 7C). In accordance with the observations above, the

variance between conditions was higher with increasing overall phase locking and strongly phase locked neurons had a significantly higher variance than weakly phase locked neurons in all cases (2-sided permutation test; $p < 0.00002$ in all cases).

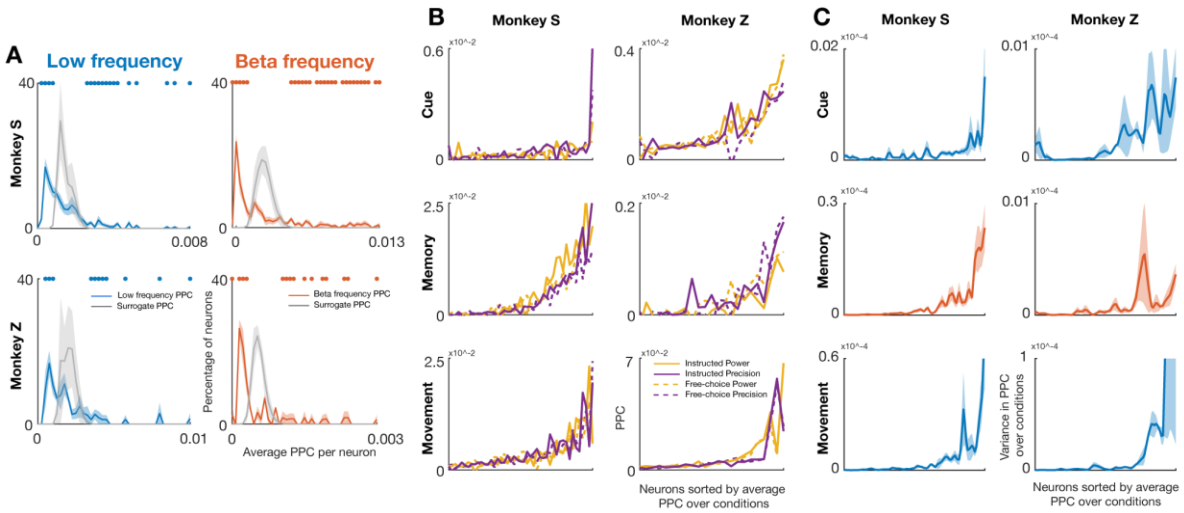

Figure 7: Strongly connected neurons in the network exhibit higher variance in PPC over task conditions. (A) Distribution of PPC values over neurons in the network compared against the average of 1000 surrogate distributions generated by distance dependent shuffling that preserves the average PPC values corresponding to within array, within cortical area and across cortical areas phase locking. **(B)** PPC per neuron for all four conditions sorted by overall PPC for exemplar recording sessions from monkeys S and Z. PPC values for the cue epoch (800ms duration windows centered from 100ms after cue onset to 500ms after cue onset) and the movement epoch (800ms duration windows centered from 300ms before movement onset to 100ms after movement onset) are from low frequency networks and the PPC values shown for the memory epoch (800ms duration windows centered from 600ms after cue onset to 1200ms after cue onset) are from beta frequency networks. **(C)** Variance across the average PPC per neuron for the four task conditions shown in (B) averaged over all recording sessions separately for monkeys S and Z. Line shadings indicate standard error across recording sessions.

LFP-site phase locking was also significantly heavy-tailed distributed (Figure S5A; network-level cluster-based permutation test, $p < 0.05$). In comparison to neurons, however, the distribution of LFP-site phase locking was less heterogeneous with more moderately and less weakly or strongly phase locked LFP-sites. Nevertheless, strongly phase-locked LFP-sites also had a significantly higher state-dependent variance than weakly phase-locked LFP-sites in all cases except for the low frequency movement period in monkey Z (Figure S5B, and S5C; 2-sided permutation test; $p < 0.05$ in all but one case).

The finding that overall strongly phase locked neurons show the strongest

condition dependent changes in phase locking in both frequencies, suggests that condition dependent reconfigurations of the network structure are driven by its strongly phase locked core of neurons.

Condition dependent changes of individual connections

Given that neurons and LFP-sites show behavior dependent changes in phase locking it remains unclear whether phase locking of individual connections in networks changes for different conditions. If condition dependent changes of individual connections are present, it would indicate that the full extent of behavior dependent changes in phase locking can only be captured at the entire network-level, without any averaging of phase locking. If, in contrast, average phase locking of neurons and LFP-sites captures all condition dependent changes, it would indicate that phase locking network analyses at the connection-level are unnecessary.

In order to capture behavior dependent changes in the entire low frequency and beta phase locking networks, we first performed principal component analysis (PCA) of all neuron-LFP pairs. PCA were performed across all conditions and time points separately for both frequency bands and per recording session. The first 30 PCs, corresponding to less than 1.5% of all neuron-LFP pairs in all cases, captured most of network-level phase locking variance in both frequencies (low frequency: 90.1 ± 0.03 % and 96.6 ± 0.01 %, and beta frequency 94.3 ± 0.01 % and 86.0 ± 0.02 % for S and Z, respectively). This result suggests that fronto-parietal low frequency and beta phase locking subnetworks are low dimensional during the performed sensory-motor transformation task and most of the meaningful network patterns are captured by these PCs.

Next, we projected low and beta frequency phase locking of one representative recording session from monkey S onto its corresponding PCs. The first two PCs of low frequency phase locking captured strong condition independent as well as strong grip-type dependent phase locking during the movement period (see Figure 8A first column for example PCs). Low frequency PCs 7 and above additionally captured

comparable amounts of condition dependent and independent phase locking during the cue period (see Figure 8A second column for example PCs). Note that cue-related phase locking captured predominantly by the higher order PCs was expected, because movement-related phase locking was much stronger and therefore dominated the earlier PCs (Figures 5B and 6A). The first PCs of beta phase locking also captured strong condition independent phase locking during the fixation and memory periods and strong condition dependent phase locking during the memory period (see Figure 8A right for example PCs). The large amount of condition dependent network phase locking captured by early PCs in both frequencies is surprising because earlier results suggest a clear dominance of condition independent phase locking (Figures 5 and 6). Therefore, these findings indicate that the full extent of condition dependent changes in phase locking can only be captured at the entire network-level.

To investigate this possibility, we estimated how much network-level phase locking per frequency over the time course of the task can be explained by individual factors neurons, LFP-sites and conditions, and how much by the interaction of these factors. Phase locking variance captured by single factors reflect global behavior dependent changes in phase locking. Phase locking variance captured by first-order factor interactions reflects fine-scale behavior dependent changes in network-level phase locking as described earlier in detail for neuron-condition and LFP-condition (Figure 6 and S3C). While phase locking variance captured by the full interaction of neurons, LFPs and conditions reflects condition dependent changes in the phase locking of individual connections in the network. To compute variance explained by the different factors and their interactions, we first linearly decomposed network-level phase locking over time into factors and factor interactions (Kobak *et al.*, 2016) separately per frequency band and recording session (see STAR Methods). Second, we calculated the variance explained by all decomposed factors. For both frequency subnetworks, neuron and LFP-site differences in phase locking explained a large percentage of phase locking variance, while average condition differences explained only a small percentage of phase locking variance (Figure 8B) in accordance with area-level results above (Figure 3B-E). The interaction factor neuron-LFP also

explained a large percentage of phase locking variance in both frequencies, suggesting the presence of subnetworks consisting of a subset of neurons and LFP-sites. As expected by the high level of heterogeneity of condition dependent phase locking of neurons and LFP-sites (Figures 6D, 7B, 7C, S3C, S5B and S5C), neuron-condition and LFP-condition interactions explained more network phase locking variance than the factor condition. However, the interaction between neurons, LFP-sites and conditions explained most condition related phase locking variance in all cases. Thus, these results suggest the presence of strong condition dependent changes in phase locking of individual connections in both frequency subnetworks.

Yet, the finding that higher order factors, especially the full interaction factor, capture more condition dependent changes in phase locking should be interpreted with some caution, because the different factors also capture different amounts of noise. The higher the order of an interaction factor, the fewer neuron-LFP pairs are averaged to estimate the factor. Hence, higher order interaction factors also capture more noise.

The surrogate data used earlier to examine condition dependent changes over time preserve the signal-to-noise ratio at all scales, but not condition dependent differences (Figure 6A; see STAR Methods). To compare condition dependent phase locking variance captured by different factors, we therefore corrected phase locking variance of all factors with the variance of the corresponding surrogate factor. Thus, the corrected variance per factor is only explained by condition dependent differences and not by any other factor or noise.

Displayed in Figure 8C is the corrected amount of network phase locking variance captured by the condition and all condition interaction factors over the time course of the task. In agreement with the results above, all factors only captured condition dependent variance in the beta band during the memory period and in the low frequency band predominantly during the movement period and for condition and neuron-condition also during the cue period. However, in contrast to uncorrected variance (Figure 8B), all four factors captured comparable amounts of corrected condition dependent variance.

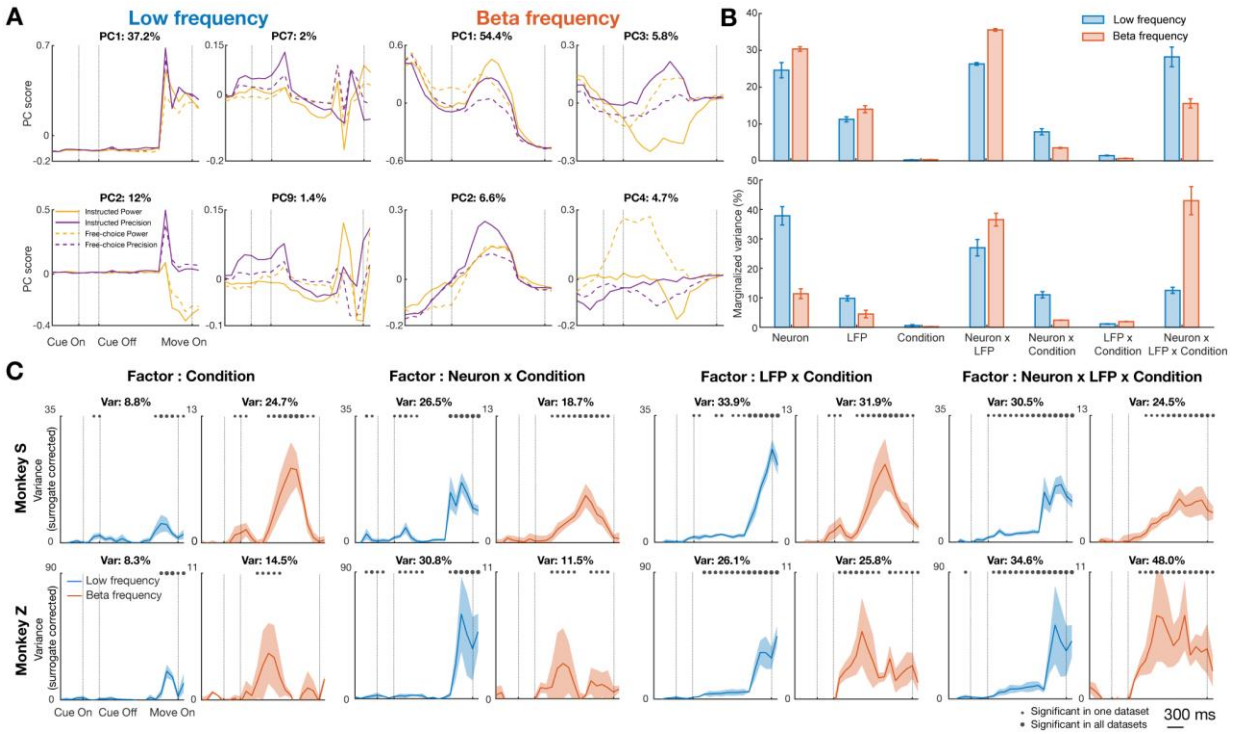

Figure 8: Network-level PCA and condition dependent variances in network factors. (A) Projection of PPC networks onto the principal components capturing highest variance during cue (PC7 and PC9 capturing 2% and 1.4% variance of low frequency networks), movement (PC1 and PC2 capturing 37.2% and 12% variance of low frequency networks) and memory (PC1, PC2, PC3 and PC4 capturing 54.4%, 6.6%, 5.8% and 4.7% variance of the beta frequency networks) epochs for an exemplar recording session from monkey S. **(B)** Percentage of the total variance captured by the factors of the network, shown separately for low frequency and beta networks. **(C)** Factor-wise variance profiles averaged over recording sessions shown separately for monkeys S and Z. Variance in the PPC values of each factor were corrected by subtracting the average and dividing by the standard deviation of the corresponding factor-wise variances computed from 1000 surrogate networks generated by shuffling condition identity before computing PPC. Dots indicate the time windows when significant condition dependent differences were detected in each illustrated factor of the low frequency and beta networks. Dot size increases proportionately with the number of recording sessions in which significant differences were detected. Line shadings indicate standard error across recording sessions. Results shown for monkey S and Z separately in (B) and (C).

Therefore, these results confirm an increasing amount of noise captured by higher order factors. Nevertheless, even for corrected variance 25.82% of all condition dependent changes in phase locking are captured by neuron-condition, 29.88% by LFP-condition, 32.46% by the full interaction neuron-LFP-condition, and only 11.82 % by the factor condition alone across both frequencies and monkeys (see Figure 8C for both frequency bands and monkeys, separately).

In summary, these results suggest that specific phase locking network patterns exist for different conditions in both frequency bands. Therefore, the full extent of behavioral changes in phase locking can only be captured at the entire

network level with neuron-resolution. At each coarser level only a small fraction of condition dependent change in the phase locking will be detected, which leads to a strong underestimation of behavior dependent differences in phase locking.

Discussion

Oscillatory synchrony in distinct frequency bands in the brain is strongly related to different cognitive and behavioral processes. For a better mechanistic understanding of how oscillatory synchronization coordinates task-dependent neuronal communication, it is essential to examine dynamic oscillatory network structure at the level of their origin: networks of neurons. In macaque monkeys performing a delayed grasping task, we estimated oscillatory network structure underlying behavioral epochs and conditions by computing pairwise phase consistency between neurons and LFPs recorded simultaneously from premotor area F5 and parietal area AIP. Neurons were preferentially synchronized either in the low or beta frequency band forming separate networks with area- and time-specificity. Low frequency synchrony was predominant during the cue and movement epochs, while beta synchrony was predominant during the fixation and memory epochs. Distinct network patterns were observed for context-types as well as grip-types of the task during the cue and memory epochs in low frequency and beta networks, respectively. In contrast, low frequency networks were distinct during the movement epoch only for grip-types, thus revealing flexible and behavior related network reconfigurations. Furthermore, a sub-group of strongly phase locked neurons contributed maximally to network reconfiguration in networks of both frequencies. These findings together might provide a framework for biologically plausible mechanistic models of flexible information processing coordinated by oscillatory synchrony.

Converging evidence supports task-specificity and site-specificity of the frequency of oscillatory synchrony (Engel and Fries, 2010). Fronto-parietal areas oscillatory synchronize in the low and beta frequency bands during tasks involving sensorimotor transformations (Scherberger, Jarvis and Andersen, 2005; Pesaran, Nelson and Andersen, 2008; Martínez-Vázquez and Gail, 2018). In congruence, behavior related oscillatory synchronization within areas F5 and AIP was almost exclusively in the low and beta frequencies during our delayed grasping task (Figures

3B and 3C). Although dominant frequency components of oscillatory synchrony can be identified at the area-level, any fine-scale differences that might be present at the level of single neurons will be lost. In the fronto-parietal network a distinct group of neurons oscillatory synchronized in the low and beta frequency bands whereas the other neurons remained non-oscillatory (Dean, Hagan and Pesaran, 2012; Dann *et al.*, 2016; Wong *et al.*, 2016). Furthermore, the oscillatory neurons were maximally inter-connected within this network, making them potential candidates for flexible information processing (Dann *et al.*, 2016). However, the dynamics of oscillatory network structure essential for flexible information processing has remained largely unexplored at the level of neurons due to methodological challenges such as the unavailability of simultaneous neural recordings or the confounding rate bias affecting spike-based measures of synchrony (Cohen and Kohn, 2011). Overcoming these challenges in our study, we examined the spatial and temporal dynamics of oscillatory networks of neurons that might be crucial for context-dependent sensory-motor transformations.

Examining oscillatory synchrony of neurons individually we identified separate groups of neurons participating in the low and beta frequency networks (Figure 4C). This implied that neurons participating strongly in beta synchrony only weakly or not at all participated in oscillatory synchrony in the low frequency band and vice versa. We wondered if area-specificity explained this separation as F5 neural averages were stronger synchronized in the low frequency band than the AIP neural averages and vice versa in the beta frequency band (Figures 3B and 3C). However, examining individual neurons in each area revealed the diversity in their preferred frequency bands. In some cases spatially neighboring neurons within an area preferred different frequencies of oscillatory synchrony. Therefore, area-specificity alone cannot sufficiently explain the separation of neural populations by their frequency of oscillatory synchrony. A plausible explanation for this strong separation is offered by studies identifying differences in oscillatory synchrony across distinct classes of neurons (Vinck *et al.*, 2013; Onorato *et al.*, 2020). We speculate that distinct neural classes with different inherently preferred frequencies might form specific

micro-circuits of oscillatory synchrony during different task-processes. In contrast to the separate groups of neurons, overlapping groups of LFPs participated in the low and beta frequency networks (Figure S3A). This finding further confirmed that the low and beta synchronized neurons were not confined to any one area since LFPs reflect averages over local neural populations. Furthermore, LFPs that can oscillatory synchronize in both frequency bands might provide a common substrate for selective entrainment of distinct neural populations in different frequencies.

Oscillatory synchrony in specific frequency bands might be a signature of certain behavioral states (Engel, Fries and Singer, 2001). During context-dependent sensorimotor transformations, a reduction in beta synchrony in motor areas has been associated with movement initiation (Canolty, Ganguly and Carmena, 2012) and increased oscillatory synchrony in the low frequency band between fronto-parietal areas has been associated with movement execution (Martínez-Vázquez and Gail, 2018). In agreement with the above studies, we found low frequency synchrony associated with active behavioral states involving interactions with the environment such as cue and movement and beta synchrony associated with steady states of the task such as fixation and memory (Figure 5B). Detecting multiple frequencies of oscillatory synchrony in the same network raised the question on the type of temporal relationship between them. Temporal relationship between oscillatory frequencies has revealed potential mechanisms for cognitive processes. For instance, simultaneous oscillatory synchronization in theta and gamma bands is proposed as a potential coding scheme for long term memorization of sequences (Lisman and Buzsáki, 2008). In contrast, a push-pull interaction with a reduction in beta synchronization triggering the onset of gamma synchronization has been shown to be crucial for the successful completion of working memory tasks (Lundqvist *et al.*, 2018). However, the temporal relation between behavior related low and beta oscillatory synchrony during sensorimotor transformations has not yet been examined at the level of single neurons. We demonstrated not only that the neurons preferentially synchronize in distinct frequencies (Figure 4C) but also that their activation times differ (Figure 5C), thus suggesting mutually exclusive low and beta

frequency oscillatory networks at the neuron-level. Oscillatory synchrony might thus facilitate switches in behavioral states by activating subnetworks of distinct, task-epoch relevant neural populations in agreement with its proposed role as a state trigger (Canolty, Ganguly and Carmena, 2012; Womelsdorf, Westendorff and Ardid, 2013).

Flexible neuronal communication over task conditions is essential for context-dependent behavior in addition to state-dependence. During movement execution low frequency oscillatory dynamics at the population level has been found to be discriminatory for task conditions (Churchland *et al.*, 2012; Elsayed *et al.*, 2016). Small but significant task condition related changes in beta oscillatory synchrony have also been reported in the population averages around planning epochs across tasks (Pesaran, Nelson and Andersen, 2008; Haegens *et al.*, 2017). However, evidence for task condition dependence of oscillatory synchrony is still lacking at the level of single neurons during sensorimotor transformations. Agreeing with earlier studies, our evaluation of neuron-level differences demonstrated significant task condition dependent variances in the low frequency networks around cue and movement epochs and in the beta networks around memory epoch (Figure 6A). Furthermore, the lower dimensional network trajectories obtained using network-level PCA were already condition-specific in the first few PCs capturing maximal variance during different behavioral epochs (Figure 8A). This finding strongly suggests task conditions to be a primary explanatory factor of variance across networks of both frequencies. Comparing condition-specific variance terms in our network demonstrated that oscillatory processes at the level of single neurons are more informative of differences across task conditions than the LFPs (Figures 8B, 6D and S3C). Intriguingly, we also found that the strongest phase locked neurons contributed maximally to changes in the oscillatory network structure revealing the heterogeneity in the neural contributions to network reconfiguration (Figure 7C). These findings agree with earlier studies showing different neurons with different oscillatory propensities (Vinck *et al.*, 2013; Dann *et al.*, 2016; Onorato *et al.*, 2020) and implies that an even smaller subnetwork within the subgroup of neurons actively participating in each

frequency band might be central to network reconfiguration. Furthermore, studies typically average over populations when examining oscillatory synchrony. However, here we showed that the lowest amount of condition-specific variance is captured in such averages. In contrast, the highest amount of variance was captured in the interaction of neurons and LFPs with conditions (Figure 8B), implying specific oscillatory network patterns underlying different task conditions. These results are the first, to our knowledge, to demonstrate larger condition-specific variances present at the level of the entire network, which can only be captured by neuron-resolved network analysis.

Overall, our findings suggest a state-dependent selection of neurons and a behavior dependent reconfiguration of networks of neurons coordinated by oscillatory synchrony during context-dependent sensorimotor transformations. These results inform future mechanistic models on flexible information processing by providing a biologically plausible framework of behavior related oscillatory synchrony.

References

Bastos, A. M., Vezoli, J. and Fries, P. (2015) 'Communication through coherence with inter-areal delays', *Current Opinion in Neurobiology*. Elsevier Ltd, 31, pp. 173–180. doi: 10.1016/j.conb.2014.11.001.

Baumann, M. A., Fluet, M. C. and Scherberger, H. (2009) 'Context-specific grasp movement representation in the macaque anterior intraparietal area', *Journal of Neuroscience*, 29(20), pp. 6436–6448. doi: 10.1523/JNEUROSCI.5479-08.2009.

Buschman, T. J. and Miller, E. K. (2007) 'Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices', *Science*, 315(5820), pp. 1860–1864. doi: 10.1126/science.1138071.

Buzsáki, G. (2010) 'Neural Syntax: Cell Assemblies, Synapsembles, and Readers', *Neuron*, 68(3), pp. 362–385. doi: 10.1016/j.neuron.2010.09.023.

Canolty, R. T., Ganguly, K. and Carmena, J. M. (2012) 'Task-Dependent Changes in Cross-Level Coupling between Single Neurons and Oscillatory Activity in Multiscale Networks', *PLoS Computational Biology*, 8(12). doi: 10.1371/journal.pcbi.1002809.

Churchland, M. M. *et al.* (2012) 'Neural population dynamics during reaching', *Nature*. Nature Publishing Group, 487(7405), pp. 51–56. doi: 10.1038/nature11129.

Cohen, M. R. and Kohn, A. (2011) 'Measuring and interpreting neuronal correlations.', *Nature neuroscience*. Nature Publishing Group, 14(7), pp. 811–819. doi: 10.1038/nn.2842.

Dann, B. *et al.* (2016) 'Uniting functional network topology and oscillations in the fronto-parietal single unit network of behaving primates', *eLife*, 5, pp. 1–26. doi: 10.7554/eLife.15719.

Dean, H. L., Hagan, M. A. and Pesaran, B. (2012) 'Only coherent spiking in posterior parietal cortex coordinates looking and reaching.', *Neuron*. Elsevier Inc., 73(4), pp. 829–41. doi: 10.1016/j.neuron.2011.12.035.

Elsayed, G. F. *et al.* (2016) 'Reorganization between preparatory and movement population responses in motor cortex', *Nature Communications*. Nature Publishing Group, 7, p. 13239.

Engel, A. K. and Fries, P. (2010) 'Beta-band oscillations—signalling the status quo?', *Current Opinion in Neurobiology*, 20(2), pp. 156–165. doi: 10.1016/j.conb.2010.02.015.

Engel, A. K., Fries, P. and Singer, W. (2001) 'Dynamic predictions: Oscillations and synchrony in top–down processing', *Nature Reviews Neuroscience*, 2(10), pp. 704–716. doi: 10.1038/35094565.

Fluet, M. C., Baumann, M. A. and Scherberger, H. (2010) 'Context-specific grasp movement representation in macaque ventral premotor cortex', *Journal of Neuroscience*, 30(45), pp. 15175–15184. doi: 10.1523/JNEUROSCI.3343-10.2010.

Fries, P. (2001) 'Modulation of Oscillatory Neuronal Synchronization by Selective Visual Attention', *Science*, 291(5508), pp. 1560–1563. doi: 10.1126/science.1055465.

Fries, P. (2015) 'Rhythms for Cognition: Communication through Coherence', *Neuron*. Elsevier Inc., 88(1), pp. 220–235. doi: 10.1016/j.neuron.2015.09.034.

Gozani, S. N. and Miller, J. P. (1994) 'Optimal discrimination and classification of neuronal action potential waveforms from multiunit, multichannel recordings using software-based linear filters', *IEEE Transactions on Biomedical Engineering*. IEEE, 41(4), pp. 358–372.

Gray, C. M. and Singer, W. (1989) 'Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex.', *Proceedings of the National Academy of Sciences of the United States of America*, 86(5), pp. 1698–1702. doi: 10.1073/pnas.86.5.1698.

Haegens, S. *et al.* (2011) 'Beta oscillations in the monkey sensorimotor network reflect somatosensory decision making', *Proceedings of the National Academy of Sciences*. National Acad Sciences, 108(26), pp. 10708–10713.

Haegens, S. *et al.* (2017) 'Beta oscillations reflect supramodal information during perceptual judgment', *Proceedings of the National Academy of Sciences*. National Acad Sciences, p. 201714633.

Kobak, D. *et al.* (2016) 'Demixed principal component analysis of neural population data', *eLife*, 5(APRIL2016), pp. 1–36. doi: 10.7554/eLife.10989.

Kraskov, A. *et al.* (2009) 'Corticospinal Neurons in Macaque Ventral Premotor Cortex with Mirror Properties: A Potential Mechanism for Action Suppression?', *Neuron*. Elsevier Ltd, 64(6), pp. 922–930. doi: 10.1016/j.neuron.2009.12.010.

Lisman, J. and Buzsáki, G. (2008) 'A Neural Coding Scheme Formed by the Combined Function of Gamma and Theta Oscillations', *Schizophrenia Bulletin*, 34(5), pp. 974–980. doi: 10.1093/schbul/sbn060.

Lundqvist, M. *et al.* (2018) 'Gamma and beta bursts during working memory readout suggest roles in its volitional control', *Nature Communications*, 9(1), p. 394. doi: 10.1038/s41467-017-02791-8.

Maris, E. and Oostenveld, R. (2007) 'Nonparametric statistical testing of EEG- and MEG-data', *Journal of Neuroscience Methods*, 164(1), pp. 177–190. doi: 10.1016/j.jneumeth.2007.03.024.

Martínez-Vázquez, P. and Gail, A. (2018) 'Directed interaction between monkey premotor and posterior parietal cortex during motor-goal retrieval from working memory', *Cerebral Cortex*. Oxford University Press, 28(5), pp. 1866–1881.

Michaels, J. A. *et al.* (2015) 'Predicting reaction time from the neural state space of the premotor and parietal grasping network', *Journal of Neuroscience*. Soc Neuroscience, 35(32), pp. 11415–11432.

Michaels, J. A. and Scherberger, H. (2018) 'Population coding of grasp and laterality-related information in the macaque fronto-parietal network', *Scientific Reports*, 8(1), pp. 1–15. doi: 10.1038/s41598-018-20051-7.

Nácher, V. *et al.* (2013) 'Coherent delta-band oscillations between cortical areas correlate with decision making', *Proceedings of the National Academy of Sciences*, p. 201314681. doi: 10.1073/pnas.1314681110.

National Research Council (2003) *Guidelines for the care and use of mammals in neuroscience and behavioral research*. National Academies Press.

von Nicolai, C. *et al.* (2014) 'Corticostriatal Coordination through Coherent Phase-Amplitude Coupling.', *Journal of Neuroscience*, 34(17), pp. 5938–5948. doi: 10.1523/JNEUROSCI.5007-13.2014.

Onorato, I. *et al.* (2020) 'A distinct class of bursting neurons with strong gamma synchronization and stimulus selectivity in monkey V1', *Neuron*. Elsevier, 105(1), pp. 180–197.

Oostenveld, R. *et al.* (2011) 'FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data', *Computational Intelligence and Neuroscience*. Edited by S. Baillet. Hindawi Publishing Corporation, 2011, p. 156869. doi: 10.1155/2011/156869.

Perkel, D. H., Gerstein, G. L. and Moore, G. P. (1967) 'Neuronal spike trains and stochastic point processes: II. Simultaneous spike trains', *Biophysical journal*. Elsevier, 7(4), pp. 419–440.

Pesaran, B. *et al.* (2002) 'Temporal structure in neuronal activity during working memory in macaque parietal cortex', *Nature Neuroscience*, 5(8), pp. 805–811. doi: 10.1038/nn890.

Pesaran, B., Nelson, M. J. and Andersen, R. A. (2008) 'Free choice activates a decision circuit between frontal and parietal cortex.', *Nature*, 453(7193), pp. 406–409. doi: 10.1038/nature06849.

Quian Quiroga, R., Nadasdy, Z. and Ben-Shaul, Y. (2004) 'Unsupervised Spike Detection and Sorting with Wavelets and Superparamagnetic Clustering', 16(8), pp. 1661–1687.

Roelfsema, P. R. *et al.* (1997) 'Visuomotor integration is associated with zero time-lag synchronization among cortical areas', *Nature*. Nature Publishing Group, 385(6612), p. 157.

Scherberger, H., Jarvis, M. R. and Andersen, R. A. (2005) 'Cortical local field potential encodes movement intentions in the posterior parietal cortex', *Neuron*, 46(2), pp. 347–354. doi: 10.1016/j.neuron.2005.03.004.

Smith, M. A. and Kohn, A. (2008) 'Spatial and temporal scales of neuronal correlation in primary visual cortex.', *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 28(48), pp. 12591–12603. doi: 10.1523/JNEUROSCI.2929-08.2008.

Vinck, M. *et al.* (2010) 'The pairwise phase consistency: A bias-free measure of rhythmic neuronal synchronization', *NeuroImage*. Elsevier Inc., 51(1), pp. 112–122. doi: 10.1016/j.neuroimage.2010.01.073.

Vinck, M. *et al.* (2012) 'Improved measures of phase-coupling between spikes and the Local Field Potential', *Journal of Computational Neuroscience*, 33(1), pp. 53–75. doi: 10.1007/s10827-011-0374-4.

Vinck, M. *et al.* (2013) 'Attentional Modulation of Cell-Class-Specific Gamma-Band Synchronization in Awake Monkey Area V4', *Neuron*. Elsevier Inc., 80(4), pp. 1077–1089. doi: 10.1016/j.neuron.2013.08.019.

Waldert, S., Lemon, R. N. and Kraskov, A. (2013) 'Influence of spiking activity on cortical local field potentials.', *J Physiol*, 591(Pt 21), pp. 5291–303. doi: 10.1113/jphysiol.2013.258228.

Womelsdorf, T., Westendorff, S. and Ardid, S. (2013) 'Subnetwork selection in deep cortical layers is mediated by beta-oscillation dependent firing.', *Frontiers in systems neuroscience*, 7(June), p. 25. doi: 10.3389/fnsys.2013.00025.

Wong, Y. T. *et al.* (2016) 'Coherent neuronal ensembles are rapidly recruited when making a look-reach decision', *Nature neuroscience*. 2016/01/11, 19(2), pp. 327–334. doi: 10.1038/nn.4210.

STAR Methods

Basic procedures

Neural activity was recorded simultaneously from many channels in two female rhesus macaque monkeys (Animals S and Z; body weight 9 and 7kg, respectively). Detailed experimental procedures have been described previously (Michaels *et al.*, 2015; Dann *et al.*, 2016). All procedures and animal care were in accordance with German and European law and were in agreement with the *Guidelines for the Care and Use of Mammals in Neuroscience and Behavioral Research* (National Research Council, 2003).

Behavioral task

Figure 2A illustrates the time course of the behavioral task described in detail previously (Michaels *et al.*, 2015; Dann *et al.*, 2016). Monkeys were trained to perform a mixed free-choice and instructed delayed grasping task. They were seated in front of a handle that could be grasped in two different ways and visual cues were displayed on a masked monitor that was superimposed on the handle using a beam splitter mirror. Trials started after the monkey placed both hands on the resting positions and fixated a red fixation disk (fixation epoch). After 600 to 1000ms, cues in the form of disks were shown next to the fixation disk for 300ms and the handle was illuminated. In the instructed context, one of two discs was displayed instructing the monkey about the required grip type (power or precision; cue epoch). In the free-choice context both disks were displayed indicating the monkey to freely choose between the two grip types. After the cue was turned off the monkey had to remain steady for a variable time of 1100 to 1500ms (memory epoch).

The turning off of the fixation light instructed the monkey to reach and grasp the target with the required grip type (movement epoch) to receive a liquid reward. Note that to encourage the monkey to perform both grip types during free-choice context, the reward was iteratively reduced every time the monkey repeatedly chose

the same grip type (mean power choice: 38.5±4.2% SD and 53.3±5.2% SD for S and Z, respectively). All trials were randomly interleaved and executed in darkness.

Electrode implantation and recordings

Surgical procedures have been described in detail previously (Michaels *et al.*, 2015; Dann *et al.*, 2016). Briefly, two floating microelectrode arrays (FMAs; Microprobes for Life Sciences; 32 electrodes; spacing between electrodes: 400µm; length: 1.5 to 7.1 mm monotonically increasing to target grey matter along the sulcus) were implanted per area in the ventral premotor cortex (area F5) and in the anterior intraparietal area (AIP), resulting in 64 electrodes per area and 128 electrodes in total. Monkeys S and Z were implanted in the left and the right hemispheres, respectively (Figure 2B). Extracellular signals from the implanted arrays were amplified and digitally stored using a 128-channel recording system (Cerebus, Blackrock Microsystems; sampling rate 30 kS/s; 0.6-7500Hz band-pass hardware filter) while the monkeys performed the delayed grasping task. All data were saved to disk and analyzed in Matlab (The Mathworks Inc., Natick, MA).

Spike detection and sorting

For spike detection, the broadband signals were first low-pass filtered with a median filter (window length 3ms) and the result subtracted from the raw signal, corresponding to a nonlinear high-pass filter. The signal was then low-pass filtered (4th order non-causal Butterworth filter, f_c: 5000 Hz). To eliminate common noise sources, principal component (PC) artifact cancellation was applied for all electrodes of each array, as described previously (Musial et al 2002; Dann 2016).To ensure that no individual channels were eliminated, PCs with any normalized coefficient greater than 0.36 (conservatively chosen) were retained. Spike waveforms were detected and semi-automatically sorted using a modified version of the offline spike sorter *Wave_clus* (Quian Quiroga, Nadasdy and Ben-Shaul, 2004; Kraskov *et al.*, 2009; Dann *et al.*, 2016).

Units were classified as single- or non-single unit based on five criteria: (1), the absence of short (1–2ms) intervals in the inter-spike interval histogram for single units; (2), the homogeneity and SD of the detected spike waveforms; (3), the separation of waveform clusters in the projection of the first 17 features (a combination for optimal discriminability of PCs, single values of the wavelet decomposition, and samples of spike waveforms) detected by *Wave_clus*; (4), the presence of well-known waveform shapes characteristics for single units; and (5), the shape of the inter-spike interval distribution.

After the semiautomatic sorting process, redetection of the different average waveforms (templates) was done to detect overlaid waveforms (Gozani and Miller, 1994; Dann *et al.*, 2016). To achieve this, filtered signals were convolved with the templates starting with the biggest waveform. Independently for each template, redetection and resorting were run automatically using a linear discriminate analysis for classification of waveforms. After spike identification, the target template was subtracted from the filtered signal of the corresponding channel to reduce artifacts during the detection of the next template. This procedure allowed us to detect spikes with a temporal overlap up to 0.2ms. Unit isolation was evaluated again, based on the five criteria mentioned above, to determine the final classification of all units into single or non-single units. Stationarity of firing rate was checked for all units and in case it was not stable over the entire recording session (more than 30% change in firing rate between the first 10 min and the last 10 min of recording) the unit was excluded from further analyses (~3% of all single units). Note that only well isolated single neurons based on the five criteria above were used for further analysis.

LFP preprocessing

To isolate LFPs, we first low-pass filtered the broadband signal with a median filter (window length 6.7 ms). Then, signals were high-pass filtered (4th order non-causal Butterworth filter, f_c: 1 Hz) and down-sampled from 30000 to 1000 samples per second by averaging every 30 consecutive samples. Broken channels and noisy

trials were identified and removed based on the total power, the Fischer z-transformed Pearson correlation coefficient of all LFP-signals, and the maximum absolute amplitude. LFP-signals in which any one of the three feature values deviated more than five standard deviations were classified as noisy and excluded from further analysis. Next, we removed the influence of the on-array ground and reference electrodes by regressing out the stratified average of all LFP-signals per array (leaving out the two highest and two lowest values per time point) of each LFP-signal.

Quantification of spike-LFP phase locking

Spike-LFP phase locking analyses were performed using custom scripts and the FieldTrip toolbox (Oostenveld *et al.*, 2011). We first binned spike events from single neurons in non-overlapping 1-ms windows to ensure that the spikes and LFP signals had the same sampling rate of 1000 Hz. Next, spikes and LFP signals of all trials were aligned to the cue and movement onsets (cue onset: -800 to 1600ms, and movement onset: -800 to 500ms). Note that we chose two alignment points because activity was locked to both events with the variable memory epoch in between.

Spike-LFP phase locking was computed using overlapping window of 800ms with a step size of 100ms separate for all four conditions. For this purpose, the instantaneous phases of LFP signals were estimated for 64 log-spaced frequencies between 2 and 130 Hz by convolving the LFP signals with Hann tapered complex sinusoids (6 cycles/frequency). For every neuron-LFP pair, we quantified the similarity in the phases of the spikes relative to the LFP signal using pairwise phase consistency (PPC), which is a measure of the strength of phase locking unbiased by firing rate (Vinck *et al.*, 2010, 2012). PPC as a function of frequency f is given by the below equation,

$$PPC(f) = \frac{\sum_{m=1}^{M}\sum_{l\neq m}^{M}\sum_{j=1}^{N_m}\sum_{k=1}^{N_l}\cos\left(\theta_{l,k}(f) - \theta_{m,j}(f)\right)}{\sum_{m=1}^{M}\sum_{l\neq m}^{M}N_m N_l} \quad (1)$$

Here, M is the total number of trials and l and m are respectively the l-th and m-th trials, j is the j-th spike of trial m and k is the k-th spike of trial l. N_m and N_l represents the number of spikes present in trials m and l, respectively. $\theta_{l,k}$ is the estimated phase of the LFP signal at the time of the k-th spike of trial l. $\theta_{m,j}$ is the estimated phase of the LFP signal at the time of the j-th spike of trial m.

In the employed version of PPC all spike-LFP phases belonging to the same trial are excluded to avoid bias due to history effects like burstiness and refractoriness (Vinck *et al.*, 2012). Because the phase locking estimates for neurons with lower spike counts have higher variance (Vinck *et al.*, 2013), we excluded PPC values estimated from neurons firing less than 70 spikes in any time window. Note that all neuron-LFP pairs on the same electrode were discarded from all further analyses to avoid artificial synchrony.

Cosine similarity analysis

We used cosine similarity analysis to compare vectors of low frequency and beta phase locking values (Figures 4C, S5C, S3A and S3B). Cosine similarity between any two vectors v_1 and v_2 is defined as

$$\cos(\theta) = \frac{v_1 . v_2}{|v_1||v_2|} \tag{2}$$

Note that the cosine similarity is equivalent to a correlation without mean subtraction and is therefore more robust for non-unimodally distributed data.

Principal component analysis (PCA) of dynamic phase locking networks

To visualize and capture behavior dependent changes in the entire low frequency and beta phase locking networks, we performed PCA of all neuron-LFP pairs across all conditions and time points separately for both frequency bands and

per recording session (Figure 8A). PCA identifies a lower number of orthogonal dimensions that explain most of the covariance of the entire phase locking networks across conditions and over time. We employed PCA to reduce the dimensionality of the matrix $ct \times nl$ to $ct \times k$ where c is the number of conditions, t the number of time windows, n the number of neurons, l the number of LFP-sites, and k the number of PCs. Note that each PC is a linear combination of all neuron-LFP pairs.

Marginalization of dynamic phase locking networks

To quantify factor-wise variances, we decomposed the low and beta phase locking networks into network factors (neuron, LFP and condition) and its interaction terms adopting the marginalization procedure employed in demixed-PCA (Kobak *et al.*, 2016). Let us denote a as the network of all phase locking values of a frequency band at a time window., a can be expressed as the sum of the uncorrelated terms as shown below

$$a = \overline{a} + \overline{a_n} + \overline{a_l} + \overline{a_c} + \overline{a_{nl}} + \overline{a_{nc}} + \overline{a_{lc}} + a_{nlc} \tag{3}$$

Individual factors in the above equation are defined as,

$$\overline{a} = \langle a \rangle_{nlc} \tag{4}$$

$$\overline{a_n} = \langle a - \overline{a} \rangle_{lc} \tag{5}$$

$$\overline{a_l} = \langle a - \overline{a} \rangle_{nc} \tag{6}$$

$$\overline{a_c} = \langle a - \overline{a} \rangle_{nl} \tag{7}$$

$$\overline{a_{nl}} = \langle a - \overline{a} - \overline{a_n} - \overline{a_l} - \overline{a_c} \rangle_c \tag{8}$$

$$\overline{a_{nc}} = \langle a - \overline{a} - \overline{a_n} - \overline{a_l} - \overline{a_c} \rangle_l \quad (9)$$

$$\overline{a_{lc}} = \langle a - \overline{a} - \overline{a_n} - \overline{a_l} - \overline{a_c} \rangle_n \quad (10)$$

$$a_{nlc} = a - \overline{a} - \overline{a_n} - \overline{a_l} - \overline{a_c} - \overline{a_{nl}} - \overline{a_{nc}} - \overline{a_{lc}} \quad (11)$$

In the above equations, a is a three dimensional matrix containing $NxLxC$ phase locking values, where N is equal to the total number of neurons, L is equal to the total number of LFP-sites and C is equal to the total number of conditions. $\overline{a}$ is the average computed over all neurons, LFPs and conditions. ⟨⟩Angular brackets denote the average over neurons (n), LFPs (l), and/or conditions (c). Furthermore, since the networks are mean centered prior to marginalization, $\overline{a} = 0$, thus reducing equation (3) to

$$a = \overline{a_n} + \overline{a_l} + \overline{a_c} + \overline{a_{nl}} + \overline{a_{nc}} + \overline{a_{lc}} + a_{nlc} \quad (12)$$

Applying equation 12 at each time window, we decomposed the low frequency and beta networks into 7 marginalized terms: neuron ($\overline{a_n}$), LFP ($\overline{a_l}$), condition ($\overline{a_c}$), neuron x LFP ($\overline{a_{nl}}$), neuron x condition ($\overline{a_{nc}}$), LFP x condition ($\overline{a_{lc}}$), and neuron x LFP x condition (a_{nlc}). After marginalizing the networks, variance was computed separately for each factor and factor interactions across all time points (Figure 8B) and separately for all time points (Figure 8C).

Cluster-based surrogate tests

We used cluster-based surrogate tests to identify significant peaks in spike-field PPC spectra and significant periods of increased phase locking variance across conditions (Maris and Oostenveld, 2007; Dann *et al.*, 2016). For the two cases, we

generated two different sets of surrogate data, while the testing procedure was the same in both cases.

The first set of surrogates was generated by randomly pairing non-simultaneous LFP trials with spike trials per condition and recomputing PPC (Figure 3C) (Perkel, Gerstein and Moore, 1967). This surrogate PPC set preserves the variability of PPC due to firing rate differences of neurons and power differences of LFP-sites over the time course of the trial per condition, while any PPC due to spike-field phase locking is removed. The testing procedure was then applied to the PPC values.

The second set of surrogates was generated by randomly reassigning spike-LFP trial pairs of the four conditions, resulting in four subsets with an equal number of trials as the four conditions (Figures 6A and 8C). Then, PPC for all neuron-LFP pairs was recomputed for the four subsets over the time course of the task. These surrogate PPC sets preserve the spike-field phase locking per neuron-LFP pair over time, but remove any condition dependent differences in phase locking. Before the testing procedure, the variance across conditions of all neuron-LFP pairs per time window was estimated for the recorded and the surrogate data. The variance was estimated for the average phase locking of neurons (Figure 6A) as well as for all marginalized factors that captured condition dependent changes in phase locking (Figure 8C). The testing procedure was then applied to the condition dependent variances.

In both cases, 1000 surrogate sets per recording session were generated and the cluster-based surrogate tests was performed in the following way:

1. z-transform every surrogate PPC value per frequency bin or surrogate variance value over time per surrogate set by subtracting the average and dividing the standard deviation of the corresponding surrogate data values.
2. Select all surrogate z-scored values greater than 2.5 and cluster them on the basis of frequency or temporal adjacency.

3. Sum clustered surrogate z-values and take the largest summed surrogate z-value.
4. Repeat steps 1-3 for all surrogate sets to construct a distribution of largest summed z-values.
5. z-transform every measured PPC value per frequency bin or variance value over time by subtracting the average and dividing the standard deviation of the corresponding surrogate data values.
6. Select all z-scored values greater than 2.5 and cluster them on the basis of frequency or temporal adjacency.
7. Sum clustered z-values and take all summed z-values.
8. For every summed z-value calculate the proportion of surrogate values that are larger than the measured summed z-values, which corresponds to the p-value.
9. Compare each p-value with a critical alpha-level (0.05 in all cases).

Note that this single comparison replaces the multiple comparisons of PPC values across frequency bins or variance values over time.

Network-level cluster-based permutation test

To test whether some neurons and LFP-sites are more or less strongly phase locked to the network than expected by chance (Figure 7A), we used a cluster-based permutation test at the network level. For this purpose, we generated 1000 surrogate networks by randomly permuting PPC values of the time and condition averaged connectivity matrix separately per frequency and recording session. Because connectivity strength decreases with distance (Smith and Kohn, 2008), we preserved distance dependent connectivity by separately permuting PPC values of neuron-LFP pairs within arrays, within areas and between areas. We next computed the average phase locking of neurons or LFP-sites of all measured and surrogate networks and binned the resulting values into 41 equally sized bins ranging from 0 to the maximum

value per frequency and monkey. For all bins, measured and surrogate data was compared by estimating t-values and clustering adjacent bin values exceeding a critical alpha-level of 0.05 separately for positive and negative t-values (corresponds to the 2.5th and 97.5th quantile of a T-distribution). Note that the performed t-tests are not the testing statistic and t-values are only used as a critical value for clustering. Subsequently, a test distribution was generated by:

1. randomly reassign measured and surrogate values of each bin while maintaining the group size.
2. re-estimate t-values and cluster adjacent bin values based on a critical alpha-level of 0.05
3. take the largest summed t-value
4. Repeat steps 1-3 1000 times

For the final statistical testing, we calculated the p-value for each measured cluster under the largest summed t-value test distribution and comparing them with a critical alpha-level of 0.05, as described above.

Permutation tests

We employed 2-sided permutation tests first, to test whether cosine similarity between low frequency and beta phase locking of all neurons and LFP-sites (Figures 4C, 5C, S3 A,B) as well as between all pairs of conditions during all three significant epochs in the corresponding frequency bands was below or above chance. Second, to test whether the condition dependent variance differences in phase locking of overall strongly and weakly phase locked neurons was below or above chance (Figure 7C, S5C). For cosine similarity testing, we generated a permutation distribution by randomly reassigning the PPC values of one frequency band or condition relative to the other frequency band or condition prior to calculating cosine similarity. For variance comparison testing, we generated a permutation distribution by randomly

reassigning condition dependent variance of overall strongly and weakly phase locked neurons (see Network-level cluster-based permutation test, Figure 7A) and calculated the difference between the group averages. In all cases, 100000 partitions were generated. Finally, the p-value of all comparisons was calculated under the corresponding random distribution and compared to a critical alpha-level of 0.05. In the case of the cosine similarity comparison between all conditions, we additionally applied Bonferroni correction for the number of condition pairs.

Supplementary Figures

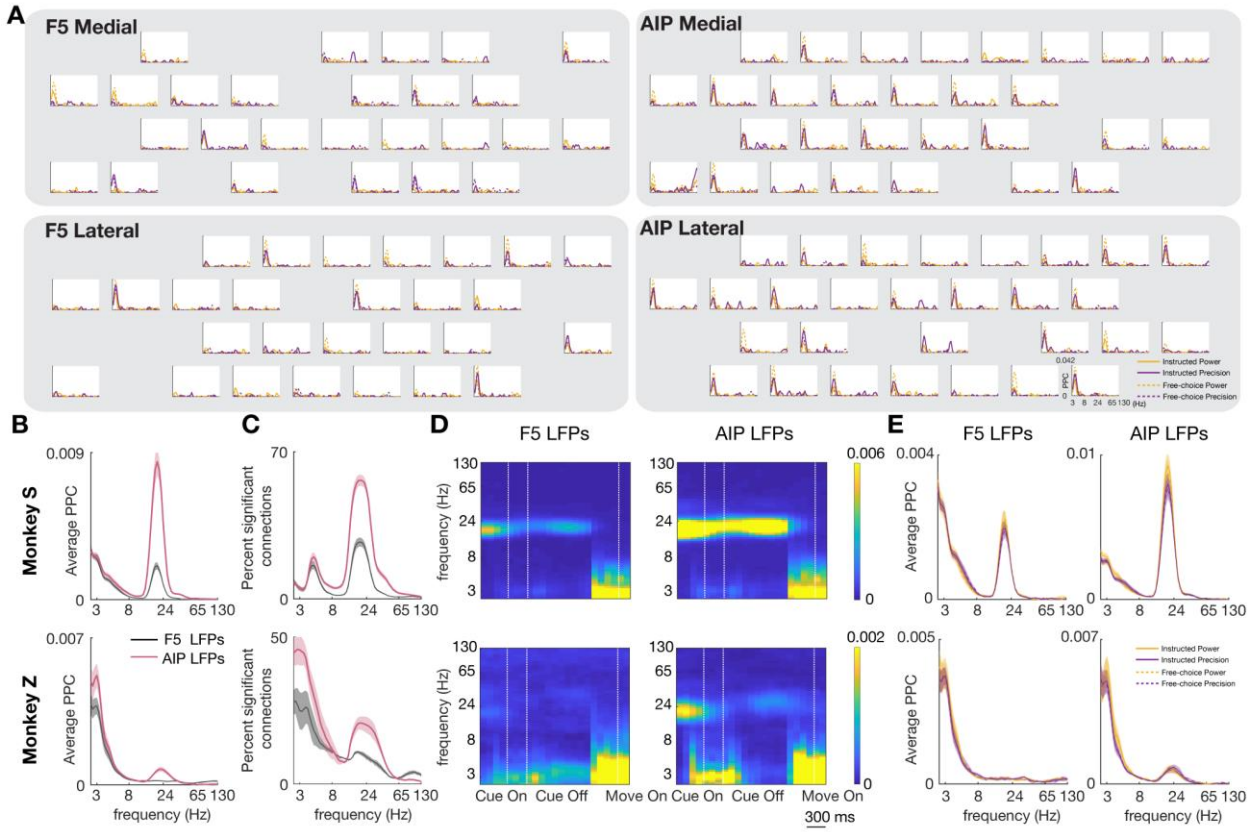

Figure S1: Example spectra and the spatial, temporal and task condition specific differences in LFP population averages. Related to Figure 3. **(A)** PPC spectra during movement epoch of a representative neuron recorded from area F5 with spatially distributed LFP signals recorded from F5 and AIP in monkey Z. **(B)** Population average phase locking in PPC spectra of LFPs recorded from areas F5 and AIP with all neurons. Line shadings indicate standard error across recording sessions. **(C)** Percentage of significant phase locking (identified using cluster-based surrogate tests, see STAR methods) of LFPs recorded from F5 and AIP with all neurons. Line shadings show standard error across recording sessions. **(D)** PPC spectrograms of LFPs recorded from F5 and AIP with neurons from both areas. In the PPC spectrograms, data were clipped at 95th percentile for visualization purposes. (in (B), (C) and (D) results are shown for monkeys S and Z separately). **(E)** PPC spectrograms of neurons recorded from F5 and AIP with LFP signals from both areas illustrating condition resolved phase locking.

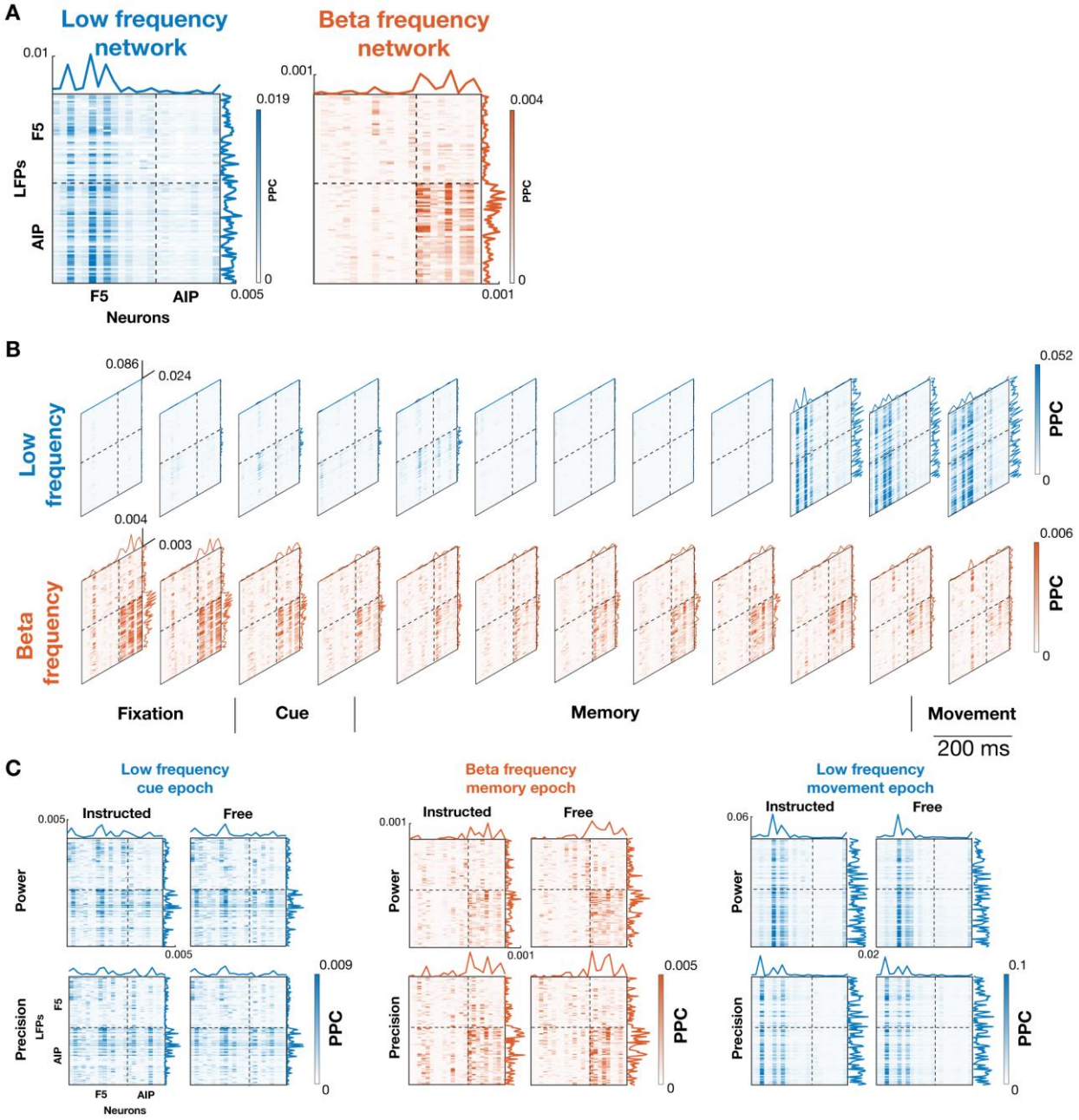

Figure S2: Low frequency and beta networks for an exemplar recording session from monkey Z. Related to Figures 4, 5 and 6. **(A)** PPCs averaged over all task conditions and time windows shown in the format of a connectivity matrix. **(B)** Network representation over time with connectivity matrices corresponding to the labelled trial epochs arranged from left to right with a temporal separation of 200ms. **(C)** Low frequency networks during the cue epoch (800ms time windows centered from 100ms after cue onset to 500ms after cue onset) and the movement epoch (800ms time windows centered from 300ms before movement onset to 100ms after movement onset) and beta frequency networks during memory epoch (800ms time windows centered from 600ms after cue onset to 1200ms after cue onset) shown for the four task conditions. PPC values in each of the network illustrations in this figure were clipped at 99th percentile for visualization purposes.

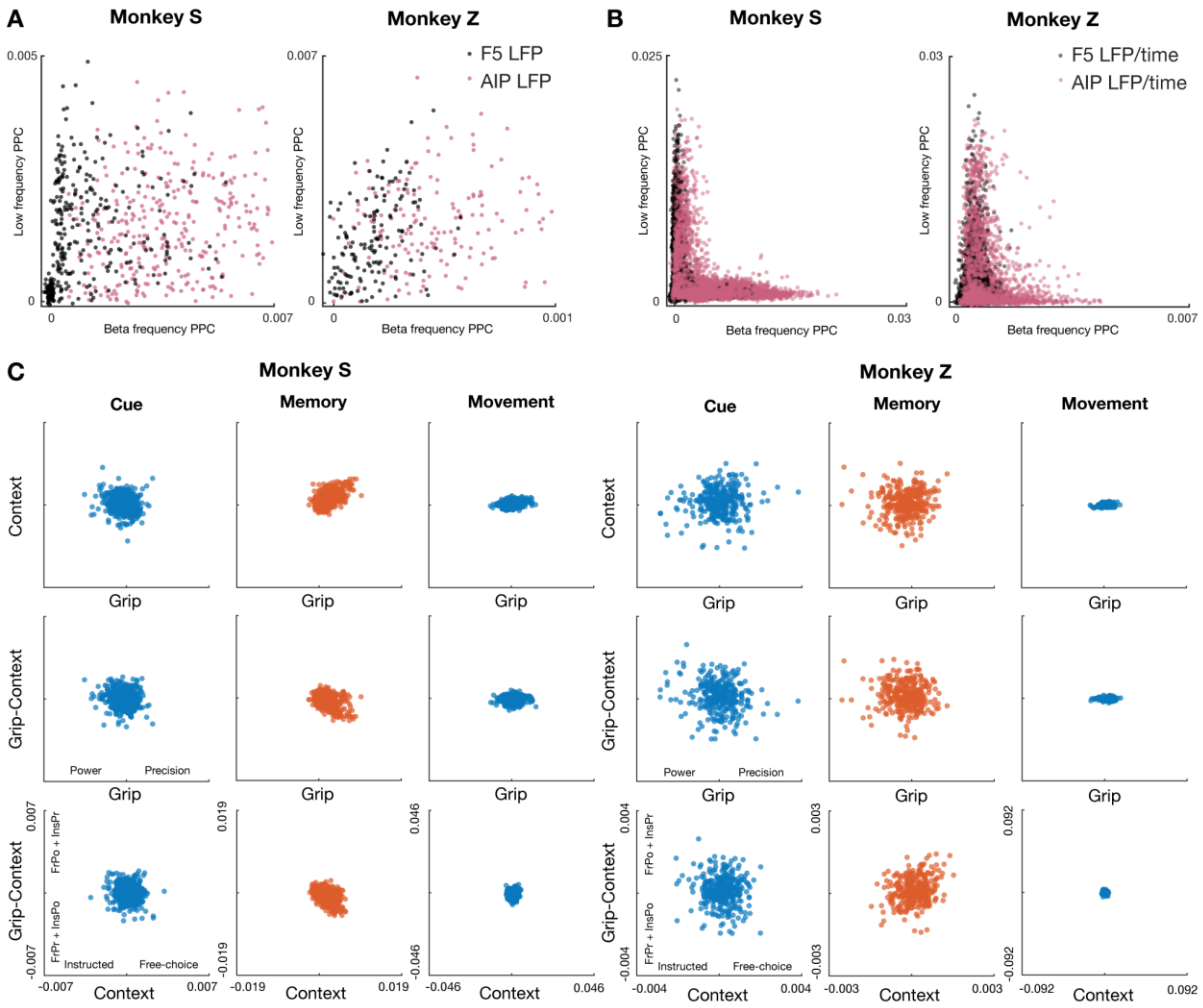

Figure S3: Groups of LFPs participating in low frequency and beta networks and their temporal dynamics. Related to Figures 4 and 5. **(A)** Comparison of average low frequency and beta PPC values corresponding to LFPs from all recording sessions. **(B)** Comparison of average low frequency and beta PPC values corresponding to LFPs at different time windows from all recording sessions. **(C)** Grip, context and grip-context interaction effects observed at the level of LFPs in the network in the low frequency band during the cue and movement epochs and in the beta frequency band during the memory epoch. Results shown for monkeys S and Z separately in (A), (B) and (C).

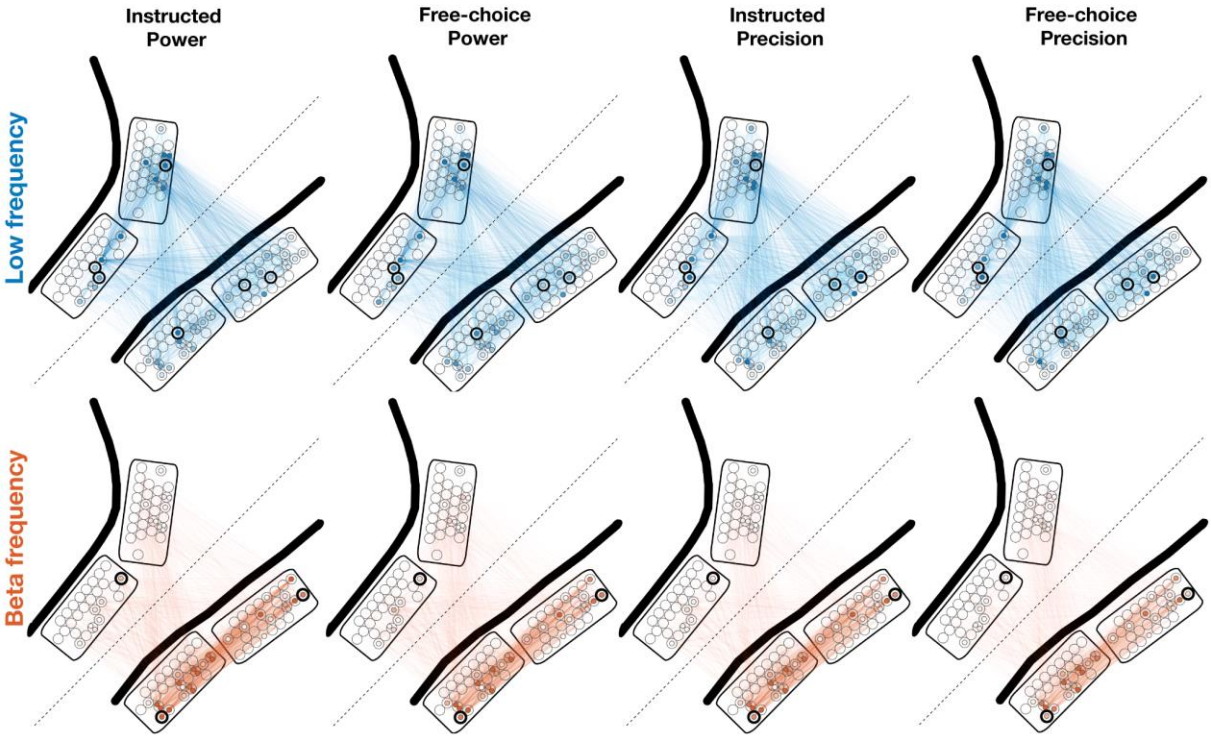

Figure S4: Differences in networks underlying task conditions illustrated in anatomical representations for a representative recording session from monkey S. Related to Figure 6. Low frequency and beta networks underlying instructed power, free power, instructed precision and free precision task conditions. Black circles drawn around select neurons to highlight the strong differences in the average PPC values they exhibit between the four task conditions.

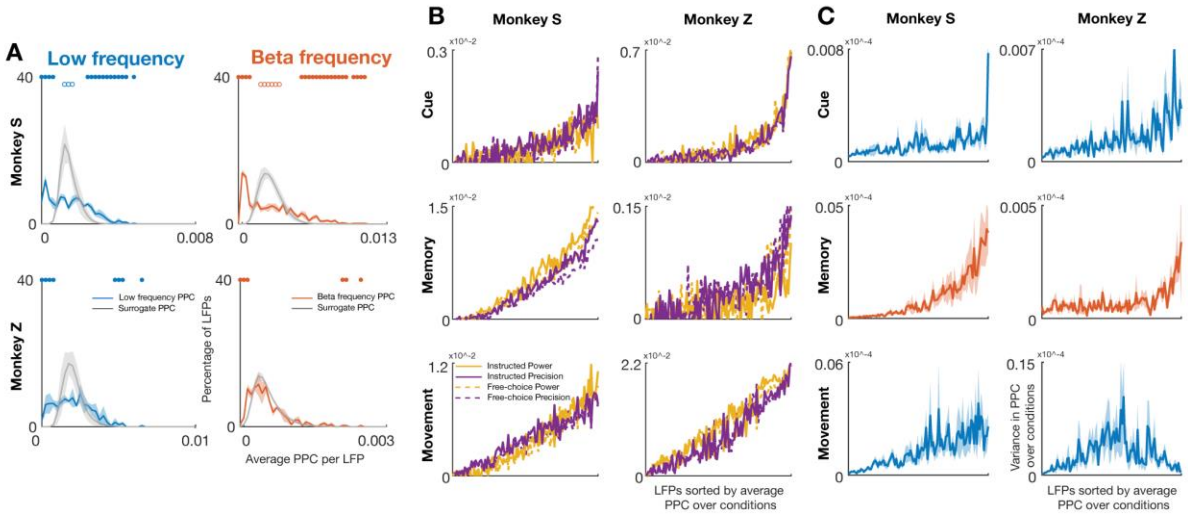

Figure S5: **Strongly connected LFPs in the network exhibit higher variance in PPC over task conditions.** Related to Figure 7. **(A)** Distribution of PPC values over LFPs in the network compared against the average of 1000 surrogate distributions generated by distance dependent shuffling that preserves the average PPC values corresponding to within array, within cortical area and across cortical areas phase locking. **(B)** PPC per LFP for all four conditions sorted by overall PPC for exemplar recording sessions from monkeys S and Z. PPC values for the cue epoch (800ms duration windows centered from 100ms after cue onset to 500ms after cue onset) and the movement epoch (800ms duration windows centered from 300ms before movement onset to 100ms after movement onset) are from low frequency networks and the PPC values shown for the memory epoch (800ms duration windows centered from 600ms after cue onset to 1200ms after cue onset) are from beta frequency networks. **(C)** Variance across the average PPC per LFP for the four task conditions shown in (B) averaged over all recording sessions separately for monkeys S and Z. Line shadings indicate standard error across recording sessions.

2.2 Precise markerless tracking of the hand during object grasping

Swathi Sheshadri[1,2*], Timo Hueser[1,3*], Benjamin Dann[1*], Hansjörg Scherberger[1,2†]

[1]German Primate Center, Kellnerweg 4, 37077 Göttingen, Germany
[2]Faculty of Biology and Psychology, University of Göttingen, 37073 Göttingen, Germany
[3]Faculty of Physics, University of Göttingen, 37073 Göttingen, Germany

[*]Equal contribution

[†]Corresponding author. Email: hscherberger@dpz.eu

Manuscript under preparation

Author contributions: S.S., T.H., B.D., and H.S. designed and planned the project. T.H., B.D., and S.S performed the experiment and built the experimental setup. S.S., and T.H. analyzed the data. S.S. wrote the manuscript. All authors revised the manuscript.

Abstract

While there are many articulated hand tracking systems, a majority of them fail in the presence of object interactions due to occlusion. However, object interactions are an integral part of studies investigating the cortical control of grasping movements. We leverage some of the state-of-the-art markerless tracking techniques (Simon *et al.*, 2017; Mathis *et al.*, 2018; Bala *et al.*, 2020) to track unrestricted hand articulations with object interactions. In contrast to the markerless tracking of whole body, absence of large-scale training datasets and occlusions during object interactions pose severe challenges to hand tracking. A previous solution that addressed these challenges required a large number of cameras ($n > 30$) mounted in specialized studios (Simon *et al.*, 2017). Our solution, on the other hand, is designed to work with as few as 5 cameras that were strategically positioned in a well-lit experimental setup to acquire behavioral videos of high signal to noise ratio. Importantly, we employ accurately annotated keypoints not only in unoccluded views but also in occluded views to train networks in DeepLabCut (DLC) (Mathis *et al.*, 2018). To this end, we 3D-reconstruct keypoints that were manually annotated in unoccluded views and reproject them onto occluded views. In addition to training networks to track occluded keypoints, this approach improves the quality of training data by exploiting geometric consistencies across multiple captured views of the same behavior. Furthermore, we incorporated a frame selection procedure based on reprojection error that allowed training networks on successively larger training datasets, which were generated automatically. We demonstrated the tracking accuracy of our system during a human grasping experiment involving fast, complex articulations with a wide range of object interactions. This solution is adaptable to macaques during grasping tasks and is a promising approach to capture natural and unconstrained hand movements.

Introduction

Our interactions with the environment largely rely on our ability to move. Hand movements in particular are crucial for object interactions that are essential to fulfil our everyday needs. Researchers have continued to address different aspects of hand movements such as sensorimotor transformations, sensori-feedback and specialized single digit movements (Schieber and Hibbard, 1993; Vargas-Irwin *et al.*, 2010; Schaffelhofer and Scherberger, 2016; Sundaram *et al.*, 2019). For the investigation of cortical control of grasping movements, it is crucial to record hand kinematic data simultaneously with neuronal recordings. To this end, the hand tracking solution must be applicable to human as well as non-human primates since insights on the neural control of movements have been predominantly acquired from the macaque animal model (Schieber and Hibbard, 1993; Lemon, 2008; Vargas-Irwin *et al.*, 2010; Dann *et al.*, 2016; Schaffelhofer and Scherberger, 2016; Michaels and Scherberger, 2018).

In macaques, hand kinematics during object interactions have been recorded using optical markers tracked by multiple cameras (Vargas-Irwin *et al.*, 2010) or sensored glove systems (Schaffelhofer and Scherberger, 2012). Optical marker systems provide accurate tracking, but face problems with occlusions. Therefore, experiments have to be designed to suit such tracking systems. This restricts the tasks being studied and makes them more unnatural such as fetching objects suspended on a swinging cable to minimize occlusions. In addition, optical marker systems require the experimenter to pre-determine the parts to be tracked and do not allow any flexibility post data acquisition. Kinematic data glove solutions based on electro-magnetic sensors solve the problem of occlusions and have been used successfully in non-human primates (Schaffelhofer and Scherberger, 2016). However, the solution requires intense animal training particularly to get them to tolerate the glove.

Markerless tracking makes tasks easier on experimental animals and facilitates faster training. Recently, with advances in deep learning, remarkable

solutions for markerless tracking such as OpenPose (Cao *et al.*, 2017), DeeperCut (Insafutdinov *et al.*, 2016), DeepLabCut (DLC) (Mathis *et al.*, 2018) and OpenMonkeyStudio (Bala *et al.*, 2020) have been developed. Among these solutions, OpenPose and DeeperCut have been employed for whole-body tracking in humans, whereas whole body tracking in monkeys has been achieved using DLC and OpenMonkeyStudio (Bala *et al.*, 2020; Berger, Agha and Gail, 2020). However, the tracked kinematics had lesser degrees of freedom than the primate hand, and did not suffer from heavy occlusions in contrast to what can be expected during real-life hand-object interactions.

Here, we propose a 5-camera markerless tracking solution that estimates hand pose during a grasping task involving reaching, grasping and a wide range of interactions carried out on a variety of objects. The cameras of the experimental setup were mounted in strategically selected locations to increase the coverage of the experimental workspace and to reduce the chances of the same keypoint being occluded in multiple camera views. From the videos acquired from the 5 cameras, we trained a ResNet (He *et al.*, 2016) based deep convolutional neural network using DLC to track 22 keypoints of the hand including finger joints, finger tips, palm and wrist. To this end, a training dataset with 1000 manual annotations was prepared. Given that the five cameras provide multiple perspectives of the same visual scene, we developed a GUI to help the human annotator make geometrically consistent annotations across all the perspectives. The GUI allowed visual inspection of images captured from the perspectives of the 5 cameras. After examining the five images, the annotator manually selected and annotated only the clearly visible keypoints in each of the images. After the first round of manual annotations, the keypoints that were manually annotated in atleast two images were 3D reconstructed and projected back to all the image planes, which were then displayed on the GUI. By making use of this information the annotator completed the annotations of all keypoints. With this technique, we reduced the possibility of making blind guesses of occluded keypoint locations and made use of geometrical constraints applicable to the multiple captured views of the same behavioral scene (Hartley and Zisserman, 2003).

Using the 1000 manual annotations, we developed an iterative network training procedure similar to techniques such as multiview bootstrapping (Simon *et al.*, 2017) and cross-view data augmentation (Bala *et al.*, 2020). Briefly, the idea is to train a network first only on manual annotations, and then automatically generate more training data from the inferences made by the trained network on behavioral videos. This procedure again exploits geometric constraints that exist across the multiple views of the same visual scene, to define a selection criterion for frames annotated by the network based on reprojection error. Frames that satisfy this criterion are added to the original training dataset with manual annotations, thereby increasing the number of examples in the training data.

By incorporating the above procedure with DLC, we selected and trained a ResNet-50 (He *et al.*, 2016) to infer unoccluded as well as occluded keypoints, thereby facilitating robust hand tracking during grasping behavior. The network had an average training error of 2.35 pixels and a testing error of 3.19 pixels. Running decoding analysis on the tracked hand kinematics revealed highly accurate (>95% on average) decoding of objects from the tracked kinematics. Furthermore, pilot experiments of 2D tracking carried out in non-human primates also revealed promising 2D tracking results. Together, these results demonstrate a hand tracking solution capable of capturing natural, unrestricted hand movements for potential applications in human as well as non-human primates.

Results

Behavioral task and video recordings

To evaluate markerless tracking, we carried out an unrestrained hand-object interaction task. The behavior consisted of a series of hand movements that included reaching, grasping, interacting with the object, and returning to a resting position (Figure 1A). Notably, the subject was completely free in choosing how to interact with the objects while staying within the experimental workspace (see Methods).

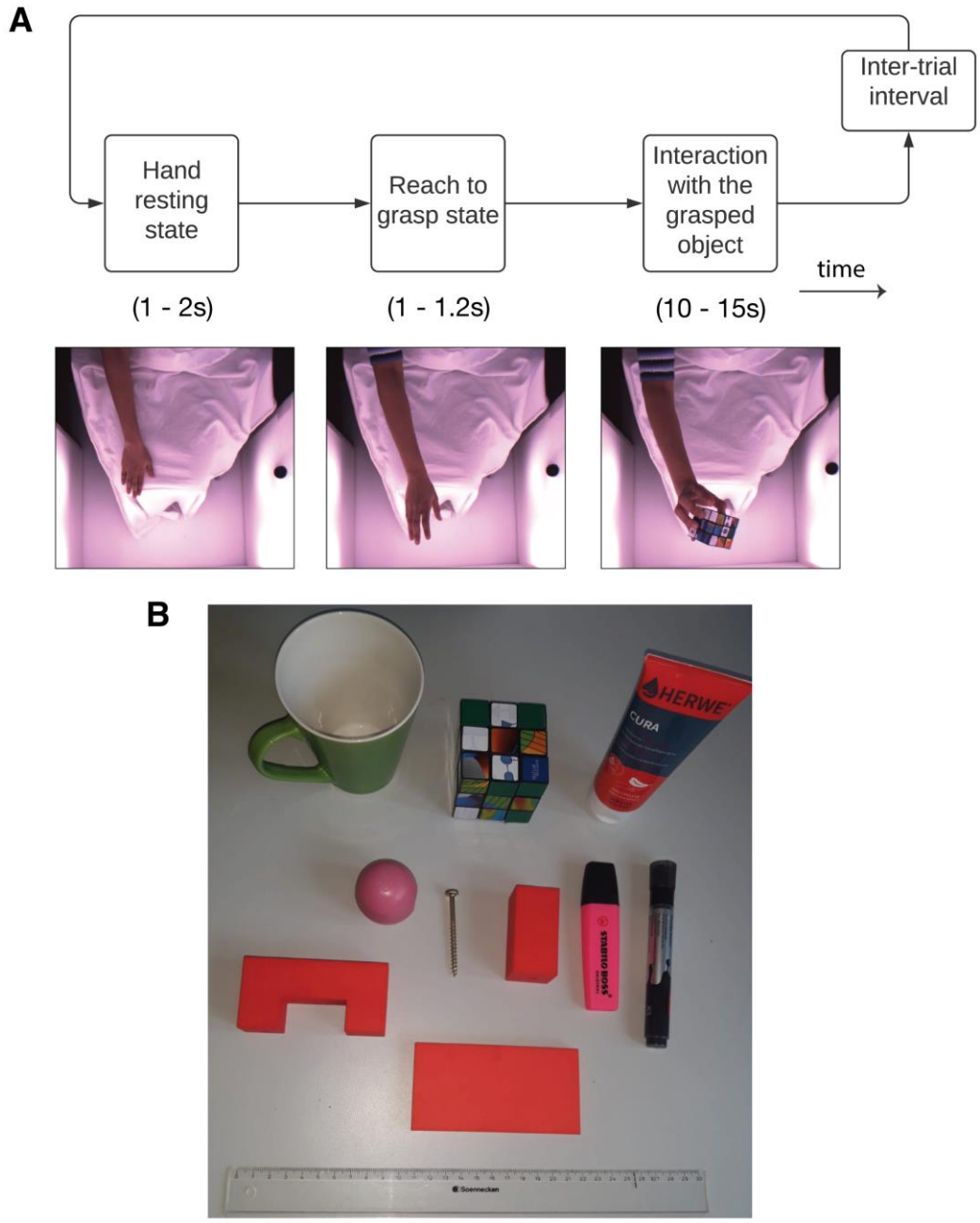

Figure 1: Unrestrained hand-object interaction task and object set. A. The trial started with the resting state of the hand (1 - 2s). Exemplar resting state is captured in the left-most photo. Hand movement to reach and grasp the object (1 – 1.2 s) followed next. Photo representative of reach-to-grasp is shown in the central illustration. Interactions with the object (10 – 15 s) was the behavioral state that lasted the longest in each trial. Exemplar object interaction behavioral state is captured in the right-most photo. At the end, following a short inter-trial interval a different object was presented in the next trial. **B.** Object set employed for the task consisting of ten different objects including a mix of everyday objects: hand cream, lip balm, marker, highlighter, nail and mug. The object set also included three geometric shapes: Rubik's cube, two cuboids of different sizes and an abstract shape.

However, to test the tracking solution on a challenging range of behaviors, she was

encouraged to employ a variety of different grasps during the interactions. 10 different objects (Figure 1B) that required different grip apertures and grip types were included in the test recording and 1 trial was carried out per object. The object set included a mix of everyday objects and geometric shapes. Hand cream tube, lip balm, marker, highlighter, nail and mug were the everyday objects that were employed for the task. Rubik's cube, cuboids of two different sizes and an abstract shape were the geometric shapes that were employed.

To track the grasping behavioral task, we built a markerless tracking system (see Methods) that included 5 cameras (Figure 2A). During the task, the participant sat in a chair facing the large aperture. The experimenter sat in front of the participant behind the large aperture and presented the objects sequentially through the aperture. Given the heavy occlusions introduced due to object interactions in the grasping task, to track the hand pose accurately from as few as 5 cameras, it is important to acquire videos that have high signal to noise ratio. Furthermore, blurring due to fast movements, saturation due to unevenly distributed light and other image artefacts must be avoided. To this end, we sufficiently lit the experimental setup with 24 LED panels that were distributed around the experimental setup, recorded behavior at high frame rate (100 Hz) and set low exposure time (1ms). In addition, the lens that we selected offered low distortion (<0.4%) even at short working distances thus largely reducing distance-dependent uneven magnification of different parts of the image. With this camera/lens configuration, even when the hand was at one of its farthest distances from the recording cameras, we could clearly resolve the features of the hand to be tracked. For example, during resting state of the hand the index finger of length ~8cm occupied ~90 pixels in the image recorded from the top-view camera. From the 5 cameras in the experimental setup, we simultaneously recorded videos lasting 2.95 minutes (17720 frames per camera) during the unrestrained grasping behavior described above.

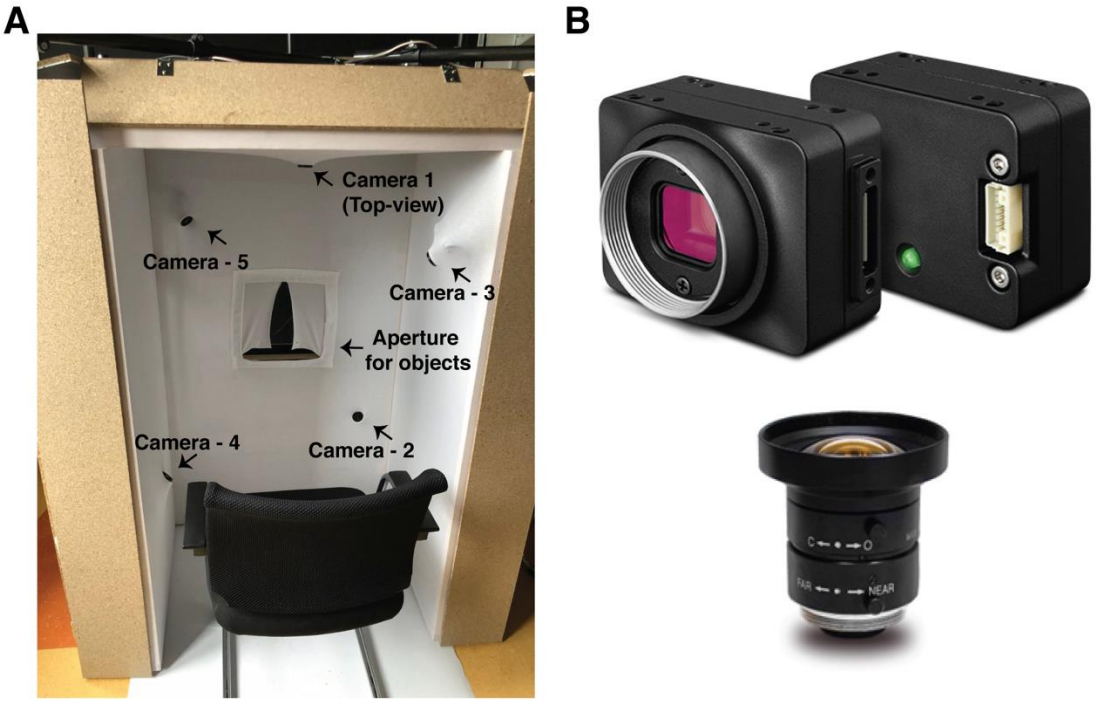

Figure 2: Experimental setup details. A. Recording setup built for markerless tracking indicating the position of the top-view, four side-view cameras, aperture for object presentation, and the participant's chair. The workspace area within the recording setup was designed to be approximately 50x50x40 cm. Five cameras positioned around the recording workspace are embedded in the white clothed frame. The top-view camera was placed at the ceiling of the frame and the side-view cameras were positioned across the walls. 24 panels of LED light were placed behind the white clothing. **B.** Chameleon camera model CM3-U3-13Y3C-CS3 from FLIR and the 3.5 mm focal length, wide angle, low distortion lens from Edmund Optics. Note that a 5 mm C-mount adapter is additionally required to connect this camera-lens pair.

To estimate hand pose during this grasping behavior we selected 22 keypoints of the hand (Figure 3). As depicted in the illustration, on each finger we tracked the tip, proximal, middle and distal joints starting from the little finger, followed by the ring, middle, index and thumb. In addition the center of the palm and the wrist were also included. To generate a deep network to learn to infer these keypoint locations in input images, a training dataset with manual annotations of the keypoints has to be created.

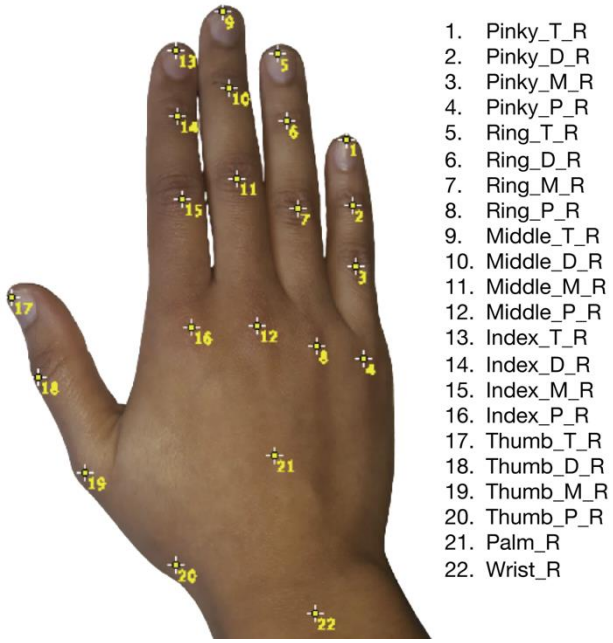

Figure 3: 22 hand keypoints for tracking. The keypoints of the hand that were tracked during the grasping task are depicted along with the corresponding label names listed on the right. In the label names T indicates tip; D indicates distal; M indicates middle; P indicates proximal; R indicates right hand.

Multiview geometry constrained annotation tool

Tracking the 22 keypoints of the hand during object interactions is not only challenging for neural networks to learn but it is also challenging for human annotators due to the presence of heavy occlusions. This makes creation of high quality training data difficult to acquire. In case of occluded keypoints, the annotator has to guess the position of occluded keypoints, thereby increasing the amount of time required for annotation and also reducing the quality of training data. Importantly, having inconsistent manual labelling across multiple views of the same keypoint can result in the network output across multiple views being inconsistent, thereby increasing the error in the 3D reconstructed (see Methods) keypoint location. To avoid this, it is crucial to enforce constraints based on multiview geometry already at the time of annotations. To this end, we developed a custom GUI that assisted the annotator in annotating occluded keypoints. The GUI visualized all five perspectives of the hand at once, allowing the annotator to visually inspect different views and

manually annotate only the clearly visible keypoints first. After the annotator provided this first set of annotations, the keypoints that were manually annotated in atleast two perspectives were 3D reconstructed and reprojected across the images displayed on the GUI. The annotator can make use of this information to complete the annotations of all keypoints across different perspectives. With this approach we not only increased the speed at which the annotations were created but also generated geometrically consistent annotations across multiple captured views of the same behavior. In relatively rare cases where a keypoint was occluded in all views, the human annotator still had to guess its location on the image based on the other visible keypoint locations. Using this GUI, we manually annotated 200 video frames from each of the five camera recordings, which added up to a total of 1000 manually annotations for network training. The 200 frames were selected by sampling uniformly from the 17720 frames from the videos recorded from each camera over the entire duration (2.95 minutes) of the behavioral task.

Iterative training and keypoints inference

For a dexterous hand movement task involving object interactions, 1000 frames may not sufficiently capture the wide variety of hand behaviors captured from multiple camera views. To improve tracking accuracy by training on a larger dataset, which better captures the diversity of data in the acquired video recordings, we adopted a procedure similar to multiview bootstrapping (Simon *et al.*, 2017) and cross-view data augmentation (Bala *et al.*, 2020). The idea exploited by these approaches is to increase the quantity of training data by including some of the inferred annotations to iteratively train deep convolutional neural networks on larger datasets. Importantly, the machine generated annotations were selected based on their conformity to the constraints of multiview geometry, thereby ensuring the quality of machine annotations (see Methods).

To infer hand keypoints in the behavioral videos, we selected a resnet-50 pretrained on ImageNet database in DLC. This network was retrained using the

above described 1000 human annotations for ~1M iterations of stochastic gradient descent (SGD). We monitored the progress in learning during the training process and stopped training when the reduction in error obtained over successive iterations was negligible. Then, the trained network was applied to infer keypoints in 2D from the video frames of the entire behavioral video recorded from the 5 cameras. For each keypoint location inferred in the 2D images, DLC additionally provides a likelihood value as output. We used the likelihood values given by DLC to select the 2D tracked coordinates to triangulate during 3D reconstruction. Only the 2D tracked coordinates associated with likelihood values exceeding a fixed threshold value of 0.9 were employed and keypoints were 3D reconstructed. Then, the 3D reconstructed data were reprojected onto the image planes of the five cameras. Following this, a subset of video frames with low reprojection error values were selected to be used as training data to train another network (see Methods). Thereby, a larger training dataset that included 7500 automatically generated training examples which were added to the initial 1000 manually created training examples was employed to train a network in DLC. Given the larger number of training examples, this network was trained for ~2M iterations of SGD. The network trained on this larger training set had an average training error of 2.35 pixels and a testing error of 3.19 pixels. In contrast, the network trained only on manually annotated frames had higher training and testing errors of 2.48 and 13.23 pixels, respectively. Therefore, the average test errors reduced by 10.04 pixels, implying a reduction of overfitting in the network trained on the combination of manually and automatically annotated images.

We used the inferences made by the network trained on the combination of manual and automatic annotations for further analysis. Exemplar annotations made by the trained network are shown for images acquired from the top-view camera (Figure 4B) and for images acquired from side-view cameras (Figure 4A). Similar to the procedure carried out in the previous iteration, only the 2D tracked coordinates associated with likelihood values exceeding 0.9 were employed for 3D reconstruction. Across the 22 keypoints that were tracked during 17720 time points of the behavioral recording, this criterion resulted in a few rare cases (386 out of 389840 keypoint 3D

reconstructions; 0.00099 %) where likelihood criterion was satisfied in less than two camera images. In such cases, we estimated 3D coordinates of the keypoints by linearly interpolating from temporally adjacent data points.

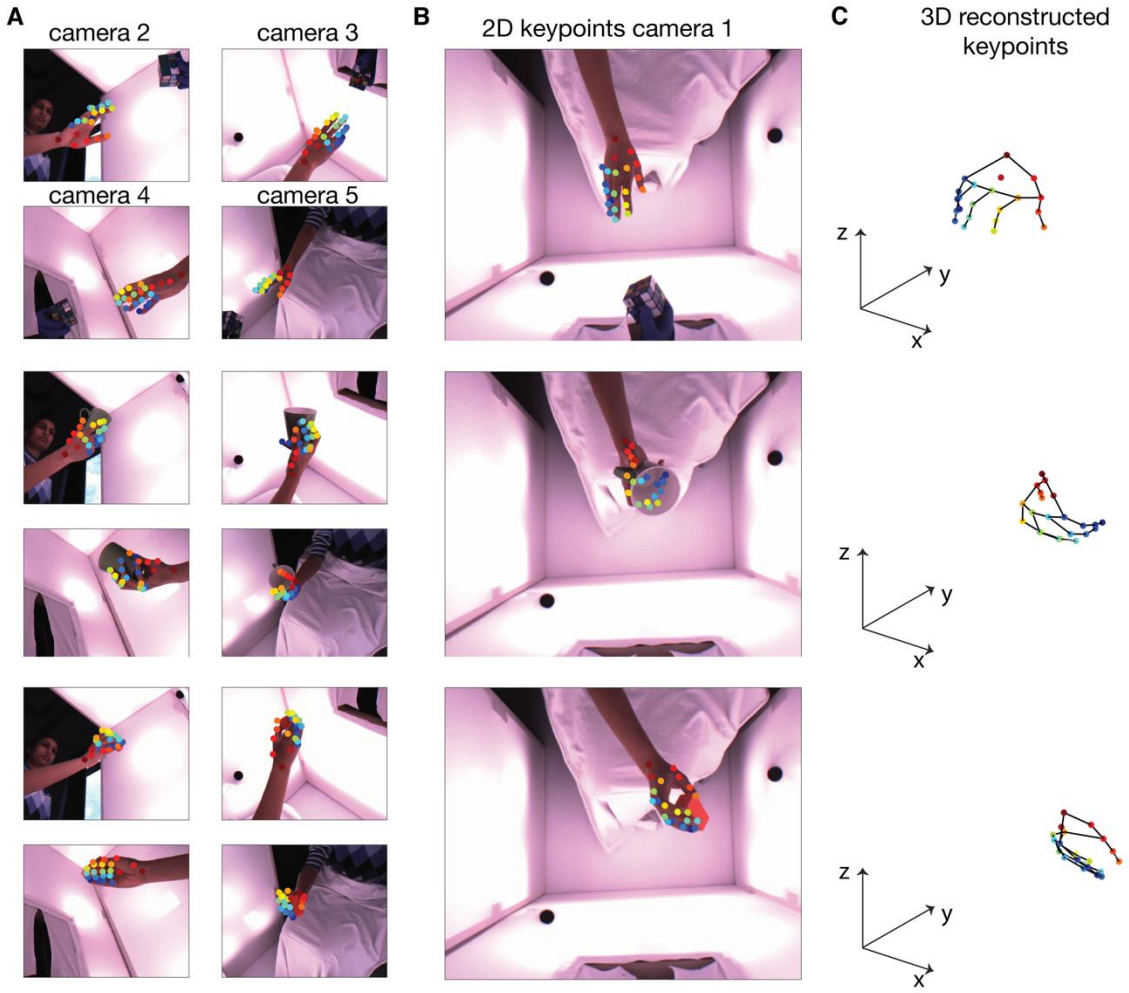

Figure 4: 3-dimesional hand pose estimation. A. Inferred positions of keypoints in 2D in video frames acquired from the four side-view cameras. **B.** Inferred positions of keypoints in 2D in video frames acquired from the top-view camera. **C.** Exemplar 3D hand-pose estimation obtained by triangulating 2D coordinates tracked on video frames recorded from five different cameras (shown in B and C). The different types of behaviors that were included during the task such as resting behavior (top), reach to grasp behavior (middle), and hand-object interactions (bottom) are shown. In A, B and C, annotations of the little finger is depicted in cool colors (shades of blue) and increasingly warmer colors are used as annotations progress through the other digits and finally the thumb is annotated in warm colors (shades of orange).

Exemplar 3D hand poses obtained by triangulating inferred 2D coordinates from the

5 camera images are illustrated (Figure 4C, also accessible in video format for the entire recording[2]). The occlusion handling capability of the proposed solution is also visible from these examples. Particularly, in the example illustrating interactions with the mug it can be clearly seen that from the perspective of the top-view camera the finger joints and the tips of almost all the fingers is occluded, yet DLC is able to infer the 2D coordinates of these keypoints.

Evaluation of grasp kinematics

For a more detailed analysis of grasp kinematics including the quantification of tracking accuracy, we employed the 3D coordinates of the hand keypoints. The keypoint coordinates tracked over the entire duration (2.95 minutes) of the behavioral recording were split into 11 time segments corresponding to the 10 grasped objects, and the resting state of the hand. First, we identified the time at which the object was grasped by visually inspecting the frames of the video recordings. Then, we used the identified time as a trigger to separate the different behavioral states in the task. 1.2 seconds of data (120 frames) preceding the object grasp time were selected as the reach to grasp behavioral state. We examined the trajectories of the 3D coordinates of the keypoints during the reach to grasp phase of the task (Figure 5). In case of poor tracking, discontinuities in the reach to grasp trajectories would be observed. Overall we obtained very smooth trajectories for the keypoints across all the ten trials, which is an indicator of well-tracked hand kinematics. To evaluate kinematic data corresponding to hand-object interactions, 10s (1000 frames) of data starting from the time the object was grasped were selected as the object interaction behavioral state. In addition, we separated 10s of resting state by combining 1s data segments sampled prior to the onset of reach to grasp corresponding to each object. To visualize and interpret the hand kinematic data during object interactions in relation of resting state, we reduced the dimensionality of the kinematic data (including 10s of interactions behavioral state per object and 10s of resting state)

[2] https://www.dropbox.com/s/qas2p45k2ug6m5b/2D_3Dmovie_10Objs_Exp.avi?dl=0

using PCA (see Methods). Prior to calculating PCA, we subtracted the 3D coordinates of the keypoint corresponding to the proximal joint of the thumb from the 3D coordinates of the other keypoints. This step was carried out to ensure PCA and decoding analysis are not dominated by the reach and positional differences of the hand.

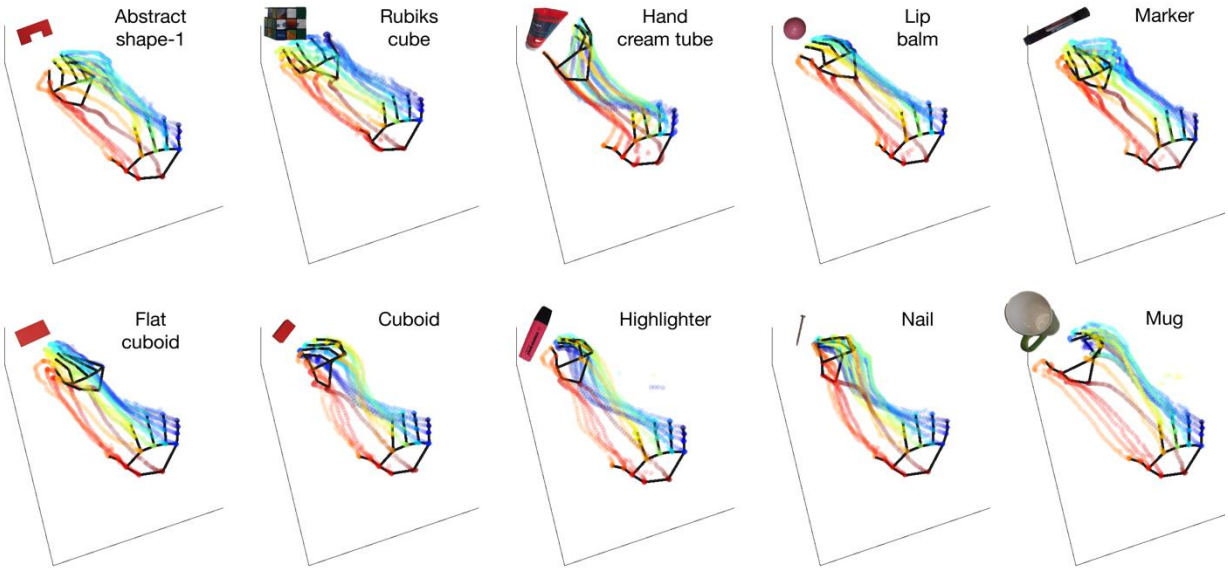

Figure 5: Single trial hand pose estimations during reach to grasp. A. Illustration of single trial trajectories of the hand keypoints during the reach and grasp behavioral states for the objects included in the object set. For each object the trial started with the resting state of the hand (bottom right in each panel) and was followed by hand movement to reach and grasp the object (~1 – 1.2 seconds) (top left in each panel).

We illustrate the projection of hand kinematic data on all possible pairs of the first six principal components (PCs) (Figure 6A). Since each object was manipulated using a large number of different grasps, we expected a strong overlap across objects in the kinematic space. In agreement with this, across the first six dominant PCs a large extent of overlap was observed in hand poses employed to grasp different objects. Interestingly, resting behavioral state that was also included in this analysis formed a distinct cluster in the low dimensional representations obtained using the first 4 PCs. Since the hand pose during the resting state is behaviorally very distinct from the hand poses employed for object interactions, these results suggest accurate tracking of hand kinematics. Furthermore, the variance across datapoints

corresponding to resting state is much smaller than that of other object interactions, highlighting the diversity of hand poses that were employed when interacting with each of the individual objects.

Next, we examined the variance captured by the PCs and found that 6 PCs were sufficient to capture 95% of hand pose variance during the entire task (Figure 6B). Lower dimensionality in hand kinematics is expected due to physiological constraints such as biomechanical coupling between some of the DOFs of the hand, which decreases the extent of independent finger movements (Buchholz, Armstrong and Goldstein, 1992; Fish and Soechting, 1992). However, to our knowledge, this is the first experiment to examine and quantify the hand kinematic dimensionality during a grasping task involving a wide-range of unconstrained hand-object interactions. Here, it is noteworthy that our object set consisted of a relatively smaller set of 10 objects and will be extended to a much larger set in future studies.

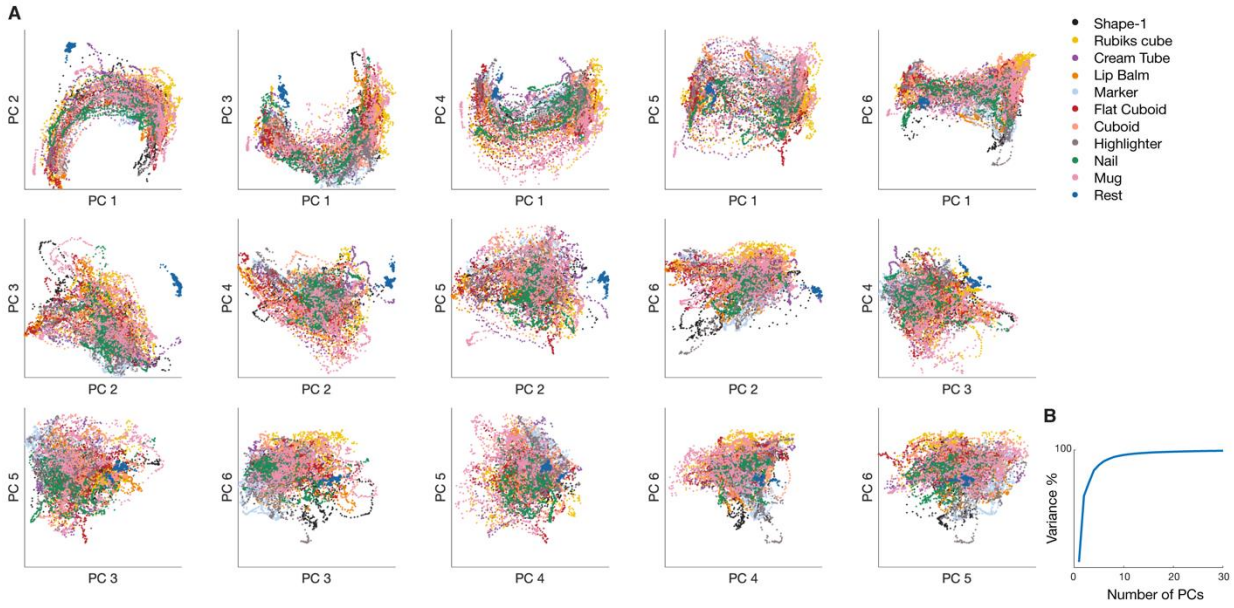

Figure 6: Low dimensional representation of hand kinematics during object interactions. **A.** Projection of 22 keypoints of the hand onto lower dimensions that capture variance in the decreasing order from PC1 (highest) to PC6. **B.** Cumulative variance explained by the principal components.

Grasp decoding from hand kinematics

We employed decoding analysis to examine whether object identity can be predicted from the tracked hand poses during object manipulation. For classification analysis, 1000 examples per object of 3D hand poses and 1000 examples corresponding to resting state were employed. First, we trained a linear discriminant analysis (LDA) classifier (see Methods) with 10-fold cross-validation. This classifier identifies linear boundaries separating the different classes of behavior. Given the high degree of overlap that was observed in the PC space, we expected some confusion between the classes of behaviors. In agreement with this, the LDA classifier performed at an average accuracy of 67.23 ± 12.35 % in classifying the 10 objects. However, as observed earlier the resting state formed a clearly separate class in the PC space and was also classified correctly with 100 percent accuracy using LDA. Interesting, in the confusion matrix (Figure 7A), it can be seen that the objects that were grasped similarly were often confused by the decoder. For example, when the true object class was marker, the decoder wrongly classified it as the highlighter in 11.4% of the cases. Similarly, when the true object class was the highlighter it was confused to be the marker frequently (24.7%). In the same way, the Rubik's cube was frequently (8.1%) confused to be the cuboid and vice-versa (10%). While the hand poses might be similar across objects, we still expected small differences to exist in the hand aperture and the overall hand articulation employed to manipulate different objects. To extract these small differences that might be missed by LDA, we employed quadratic discriminant analysis (QDA), which identifies non-linear boundaries of separation between different classes. The accuracy of classification improved with the QDA classifier reaching average accuracies of 95.6 ± 2.49 over the ten object classes. Similar to the LDA classifier, QDA classifier had very high accuracies in decoding the resting state (98.5%) and was found to most frequently confuse similarly grasped objects such as highlighter and marker (2.7% of the cases highlighter wrongly classified as marker, 1.6% vice versa), and Rubik's cube and cuboid (1.5% of the cases Rubik's cube wrongly classified as cuboid and 1.7% vice versa).

Figure 7: Decoding grasped objects. A. Confusion matrix illustrating the percentage of correct (green shades) and wrong classifications (red shades) for each object obtained using LDA. The rows correspond to the true class and the columns correspond to the predicted class of the objects. Entries on the principal diagonal depict correct classifications and off-diagonal entries depict incorrect classifications. **B.** Same as A but for classifications made using QDA.

Adaptability of the proposed solution to macaques

We tested the 2D tracking component of our markerless tracking solution that relies on DLC in macaques during a turntable task (Figure 8). For this pilot experiment behavioral videos were captured from a single camera (GoPro Hero4). During the task, a macaque monkey was presented objects on a turn table one by one and was rewarded with fruits for successfully reaching, grasping and lifting the presented object. We trained the resnet-50 based architecture in DLC using 200 manually labelled examples. We used the same set of keypoints for the study as was used in the human experiment (Figure 3). However, additionally the elbow and shoulder was tracked in this experiment. By training over 300000 iterations of stochastic gradient descent algorithm, the average training error reduced to 1.35 pixels and testing error to 5.93 pixels, which is smaller than the diameter of the filled markers used to annotate the keypoints (see Figure 8). These preliminary results demonstrate the scalability of the proposed DLC based solutions to macaques.

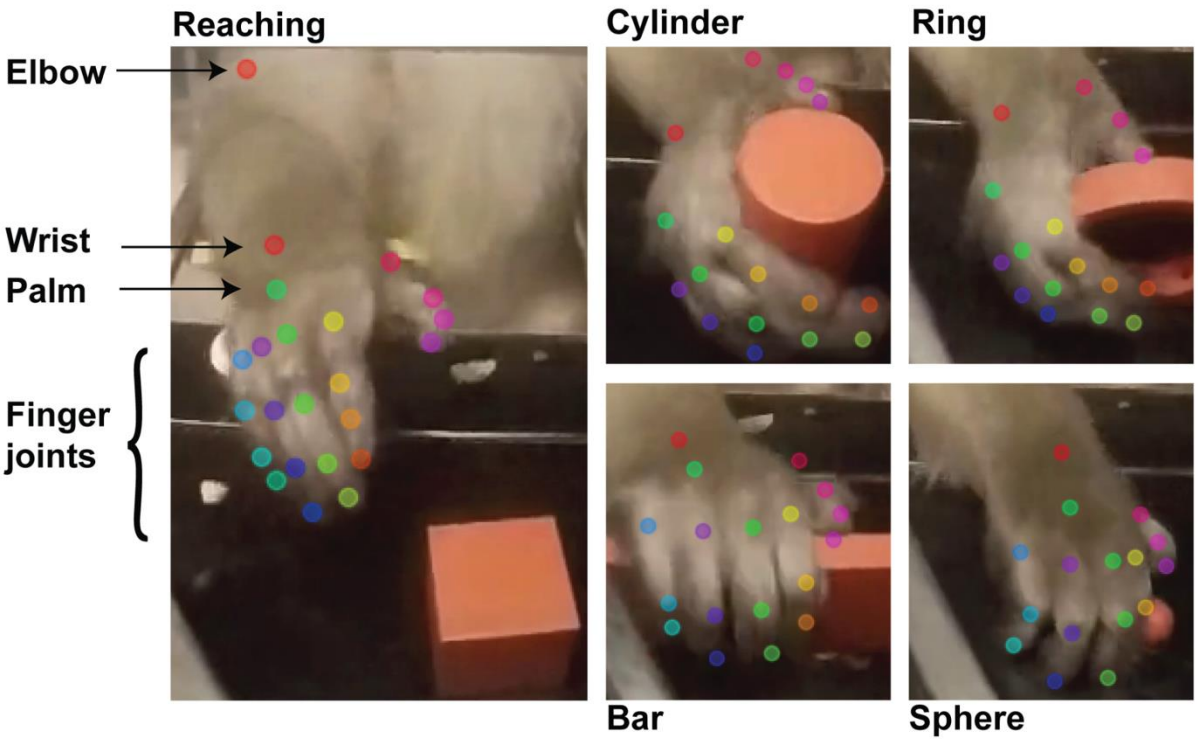

Figure 8: Example keypoint inferences made by DLC-trained network on macaque hand. Example inferences made by a DLC-trained network illustrating markerless tracking of the proximal, middle and distal joints, and tips of each finger, back of the palm, wrist and elbow joints of right hand of a macaque while the macaque grasped objects presented on a turn table.

Overall, we demonstrated accurate tracking of hand kinematics during natural, unrestricted hand movements involving a wide range of object interactions, which was implemented while employing a relatively economical experimental setup.

Discussion

Summary

In this paper, we presented the developmental and algorithmic details of a markerless grasp tracking paradigm. The system is capable of tracking hand keypoints from video frames of unconstrained hand movements with object interactions acquired simultaneously from 5 cameras. One of the main challenges for hand tracking, especially during object interactions, is the problem of occlusions. We

addressed this by generating geometrically consistent manual annotations of the unoccluded as well as occluded keypoints that were then used to train a deep convolutional neural network to track hand kinematics. The other main challenge for hand tracking is the absence of large-scale training datasets. To address this, we programmed a technique that automatically added machine-generated annotations to the set of annotations generated by the human annotator. First, the network trained on manual annotations was employed to predict keypoints on the acquired behavioral videos. Then, by exploiting geometric constraints existing across the captured multiple views of the visual scene, the machine generated annotations were validated. In this process, a subset of machine annotated frames that had low reprojection errors for all keypoints were selected and added to the training dataset.

The solution proposed here was validated by tracking hand kinematics during a grasping task with many degrees of freedom and high levels of object interactions. The results from the validation experiment clearly demonstrated the technical capabilities of the solution. Importantly, the proposed markerless tracking system allows recording natural and unconstrained hand kinematics. This makes it a preferable approach in lab experiments aimed to understand how grasping movements are encoded by the brain.

Advantages of the proposed markerless tracking system

There are several advantages of markerless tracking of which some are generally applicable to most vision-based tracking systems and some are specific to the solution proposed here. First, markerless tracking systems are completely contactless and do not interfere with hand movements thereby capturing natural hand movements. Second, the camera-based systems do not interfere with other signals such as electrophysiological signals that might be simultaneously acquired during the experiments. Third, depending on experimental requirements, the frame rate, magnification and resolution can be scaled up or down easily by selecting the cameras and lens that meet the technical requirements. Fourth, the keypoints to be

tracked can also be altered post data acquisition, in contrast to other instrumented gloves (Schaffelhofer and Scherberger, 2012) and vision-based systems requiring sensor locations or markers to be pre-determined before data acquisition (Vargas-Irwin *et al.*, 2010). Fifth, in contrast to vision based systems that fail in the presence of occlusions, here we take advantage of the multiple captured views to estimate keypoints in occluded views from unoccluded views of the same keypoint.

Improving automatic annotation augmentation technique

Using the multiview annotation augmentation method described in this paper, we gained 7500 automatically annotated frames from a total of 88600 frames recorded from all 5 cameras. Interestingly, the number of network annotated frames that we gained were unevenly distributed across the five cameras. We obtained 2193 additionally annotated frames from the top-view camera, 152, 1511, 138 and 3506 frames from side-view cameras numbered 2, 3, 4, and 5, respectively. It is surprising that from two of the cameras, we have much lower number of frames that were selected for training than the other three. One of the reasons for this could be because they exhibit poor 2D tracking for a small subset of keypoints. This implies that there might still be some keypoints in these camera perspectives that are useful for training purposes. Furthermore, there still exists the possibility of estimating the poorly predicted keypoints in an individual camera's perspective from the other cameras. This has already been implemented in the GUI-based manual annotation tool described earlier in the paper and can be applied during the process of automatic augmentation of annotations as well. To detect and remove outliers prior to 3D reconstruction, we are developing a keypoint-wise outlier rejection criterion to exclude camera perspectives that poorly track a particular keypoint. The idea is to perform keypoint-wise 3D reconstruction from all possible pairs of cameras and examine the distances between the 3D reconstructed points. In case the keypoint is tracked relatively poorly in any one of the perspectives, then the 3D reconstructions that were made using this camera will form a separate and distant cluster. Therefore,

the distances between the 3D reconstructed keypoint coordinates can be employed to identify outlier camera views for each keypoint. The poorly tracked keypoint can then be replaced by reprojecting the 3D reconstructed keypoint estimate obtained from the other perspectives. This improves not only the quality of 3D reconstruction but also allows an automatic reannotation of 2D coordinates in the rejected camera perspective. By including this in our future implementations, we expect not only a more evenly distributed training dataset across camera perspectives but also an overall increase in the number of selected frames that can be used for network training.

Adding musculoskeletal model

The keypoints (in the context of OpenSim models referred to as markers) that have been selected and tracked in this study can further be employed to estimate muscle-tendon and skeletal kinematics by fitting an OpenSim model (Delp *et al.*, 2007). To this end, we plan to use a primate-specific model of the upper arm and hand, that provides joint angles of 27 degrees of freedom and the muscle lengths of 50 musculotendon units (Schaffelhofer and Scherberger, 2012). While the generic model includes a set of predefined markers, the model must be adapted for use across different hands to accurately run simulations. The generic model can be adapted to fit different hands by using the scaling tool, which is available in the OpenSim software platform. The scaling tool requires the experimental markers to be tracked during the static pose of the hand. Using this data, anatomical segments of the generic model are updated ensuring a close match between the experimental markers and the markers defined in the model. After scaling the model, experimental marker kinematics tracked during the entire behavioral task can be imported to OpenSim. Following this, using the 'run inverse kinematics' tool, the marker positions can be converted to joint angles and muscle length vectors. In addition to transforming data from marker position-domain to joint-angles and muscle-length domains, this model fitting procedure might also refine tracking further by imposing physiological

constraints, thereby resulting in biologically feasible hand poses.

Computationally more efficient training

The solution proposed here requires multiple network training iterations and is therefore computationally expensive. To improve the computational efficiency of the proposed solution, the following pre-processing step can be adapted. Prior to training networks to detect keypoints of the hand, the video frames can be cropped to retain only the part of the image containing the hand. Faster region based convolutional neural networks (R-CNNs) (Ren *et al.*, 2015), RetinaNet (Lin *et al.*, 2017) are among the leading solutions that are existing to identify region of interest in images. These solutions employ ConvNets, such as MobileNet (Howard *et al.*, 2017) and ResNet (He *et al.*, 2016) that take images as inputs, identify bounding boxes that capture the objects in the input images and then classify the objects in the images. From the bounding box outputs generated by these networks, the bounding box that captures the hand can be extracted. Adding such a region of interest detector prior to feature tracking can substantially boost the computational speed our the solution we proposed. However, to accurately infer hand region on images captured from specific experimental setups, the networks might require additional retraining on images captured during the experiment.

Methods

Hand movement video acquisition system

We designed a five-camera system (Figure 2A) to acquire video data for hand pose estimation during the unrestrained hand-object interaction task. The cameras were mounted onto an external frame that was built using carbon fiber tubes. Adjustable arms were used to select and stabilize the height and orientation of the cameras to effectively cover the workspace, which spanned a volume of 50x50x40 cm

around the center of the clothed frame. 24 LED light panels (LM8 100BI bi-color on-camera LED lamp) were fixed around the frame and placed behind the white cloth to sufficiently and uniformly light the workspace while avoiding undesirable saturation effects in the recorded videos. Apertures were created in the cloth for the recording cameras and for object presentation during the experiment. Machine vision cameras (Chameleon, FLIR) were employed in combination with a 3.5 mm focal length, low distortion (<0.4%) lens (Edmund optics) (Figure 2B). The cameras were equipped with a ½" CMOS sensor, 4.8μm pixel size, and a global shutter. The sensors operated at a quantum efficiency of >40% (for all three color streams), making them acceptable for use even under low lighting conditions and/or low exposure times that were employed in our experiment to avoid blurring due to fast movements. Furthermore, the availability of Spinnaker SDK, which is an API built to support machine vision cameras from FLIR, made the customization of camera settings programmable. We programmed the cameras to stream videos with a resolution of 1,280 x 1,024 pixels, in BayerRG format (12 bits per pixel), at 100 frames per second. To accurately reconstruct hand pose in 3D, it is essential that the cameras are synchronized and capture the same static scene from different perspectives. To this end, we employed a 3.3 V TTL pulse generated from an external microcontroller (Arduino Uno). This pulse was supplied to the cameras via the General Purpose Input/Output (GPIO) ports and triggered the shutters of the cameras to capture images of the scene from different camera perspectives simultaneously. Each camera streamed videos (~195 MB per second, JPEG lossless compression) to a recording computer (Precision 3930 Rack XCTO Intel Core i7-9700, 8 Core, 12MB Cache, 3.0Ghz, 4.8 GHz Turbo w/UHD Graphics 630) via USB port. One of the five recording computers was designated as the master, and it received video data from the remaining computers via a network connection and saved the behavioral videos to a local disk.

DeepLabCut network training

All data analysis was performed using DeepLabCut (DLC), pose3d, customized

Matlab (The Mathworks Inc., Natick, MA) and python codes. We used Lambda Quad RTX (Lambda Labs) workstation with four NVIDIA GeForce RTX 2080 Ti to train deep neural networks to track keypoints on the hand (Figure 3). DLC with Resnet-50 pretrained on the ImageNet database (Deng *et al.*, 2009) was used for training. DLC uses an adapted version of ResNet architecture that was also employed in DeeperCut (Insafutdinov *et al.*, 2016). In the adapted architecture, the softmax output layer employed in ResNet is replaced by deconvolutional layers. The number of deconvolutional layers matches the number of keypoints to be tracked. The network output is given by the activation of each deconvolutional layer, which is a heatmap that represents the likelihood of the position of the corresponding keypoint across all pixels of the image. The peak of the heat map corresponds to the position of the keypoint with the highest likelihood value. To train the network on the hand tracking task, we selected a training dataset that was selected uniformly over time. The keypoints were then manually annotated in all the images of the training dataset using a customized annotation GUI tool. Then, the dataset was randomly split into training (90%) and testing (10%) datasets. On the training dataset, stochastic gradient descent (SGD) optimization procedure with a batch size of 1 and cross-entropy loss function was employed for training. Following network training, testing and training errors were quantified using the Euclidean distance measure between the human-annotated and network-inferred annotations. The Euclidean distances were computed for each keypoint separately and averaged over all keypoints and image frames.

3D reconstruction

To reconstruct the 3-dimensional coordinates of the keypoints tracked in 2D across multiple cameras, we employed pose3d (Sheshadri *et al.*, 2019). The first step in pose3d is camera calibration. For this, we first recorded videos of a checkerboard of known square size from the five cameras. Checkboard videos were captured simultaneously from the top-view (primary) and each of the side-view (secondary)

cameras sequentially (stereo camera calibration). From the checkerboard recordings, the extrinsic and intrinsic parameters of the camera, including the focal length, the distortion in the camera image and the location of the cameras (mounted around the side-walls of the setup) with respect to the top-view camera were estimated. Using the calibrated camera parameters, pose3d carried out triangulation to estimate the 3D coordinates of the keypoints. For 3D reconstruction, only those 2D coordinates that were inferred by the trained network with likelihood values exceeding 0.9 were employed. 2D coordinates of the keypoint from a minimum of 2 cameras are required for triangulation. When more than 2 cameras captured the keypoint a least-squares estimate of the keypoint was calculated.

Automatic annotation augmentation

From the first iteration of training carried out using human-made annotations, we automatically generated more training data to improve the accuracy of hand pose estimation as follows. First, the predicted 2D keypoints across the 5 cameras that exceeded the likelihood threshold (0.9) were triangulated to estimate 3D coordinates of the keypoints. Second, the 3D reconstructed data were reprojected onto the image planes of the 5 cameras. Third, the frames in which the differences between the reprojected and DLC predicted keypoints were less than 10 pixels for each keypoint were selected as additional frames. In other words, the selected frames had all keypoints annotated with a likelihood value > 0.9 and had reprojection errors < 10 pixel per keypoint. Fourth, the network training procedure in DLC was carried out again using the combination of manual and automatically generated keypoint annotations.

Principal component analysis

We used principal component analysis (PCA) to reduce the dimensionality of the hand kinematic data (keypoints x timepoints) and to quantify the number of

dimensions required to capture large variance components in the data (Figure 6). For PCA, keypoints were set as variables and time points corresponding to the ten grasped objects and the resting position of the hand were used as examples. PCA identifies orthogonal dimensions that explain most of the covariance in the hand kinematic data. The PCs are numbered in accordance with the amount of variance captured by them, with the first PC capturing the highest amount of variance.

Decoding analysis

To decode objects from kinematic data, we used discriminant analysis. Discriminant analysis is a supervised classification procedure that maximizes the between-class variance in relation to the within-class variance. While linear discriminant analysis (LDA) allows linear separation of data belonging to different classes, quadratic discriminant analysis (QDA) allows non-linear separation of data. Decoding accuracy of discriminant analysis was quantified by 10-fold cross validation procedure as follows. First, we randomly created 10 subsets of data. Second, for each subset of data, the remaining 9 subsets were used to train the model for discriminant analysis on the ten objects, rest classification task. The held-out subset, referred to as the test subset, was then employed to calculate classification accuracy.

References

Bala, P. *et al.* (2020) 'OpenMonkeyStudio: Automated Markerless Pose Estimation in Freely Moving Macaques', *Nature Communications*. Springer US, (2020), pp. 1–12. doi: 10.1101/2020.01.31.928861.

Berger, M., Agha, N. and Gail, A. (2020) 'Wireless recording from unrestrained monkeys reveals motor goal encoding beyond immediate reach in frontoparietal cortex', *eLife*, 9, pp. 1–29. doi: 10.1101/305334.

Buchholz, B., Armstrong, T. J. and Goldstein, S. A. (1992) 'Anthropometric data for describing the kinematics of the human hand', *Ergonomics*, 35(3), pp. 261–273. doi: 10.1080/00140139208967812.

Cao, Z. *et al.* (2017) 'Realtime multi-person 2D pose estimation using part affinity fields', *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017-Janua(Xxx), pp. 1302–1310. doi: 10.1109/CVPR.2017.143.

Dann, B. *et al.* (2016) 'Uniting functional network topology and oscillations in the fronto-parietal single unit network of behaving primates', *eLife*, 5, pp. 1–26. doi: 10.7554/eLife.15719.

Delp, S. L. *et al.* (2007) 'OpenSim: open-source software to create and analyze dynamic simulations of movement', *IEEE transactions on biomedical engineering*. IEEE, 54(11), pp. 1940–1950.

Deng, J. *et al.* (2009) 'Imagenet: A large-scale hierarchical image database', in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, pp. 248–255.

Fish, J. and Soechting, J. F. (1992) 'Synergistic finger movements in a skilled motor task', *Experimental Brain Research*. Springer, 91(2), pp. 327–334.

Hartley, R. and Zisserman, A. (2003) *Multiple view geometry in computer vision*. Cambridge university press.

He, K. *et al.* (2016) 'Deep residual learning for image recognition', in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778.

Howard, A. G. *et al.* (2017) 'MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications'. Available at: http://arxiv.org/abs/1704.04861.

Insafutdinov, E. *et al.* (2016) 'Deepercut: A deeper, stronger, and faster multi-person pose estimation model', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9910 LNCS, pp. 34–50. doi: 10.1007/978-3-319-46466-4_3.

Lemon, R. N. (2008) 'Descending pathways in motor control', *Annual Review of Neuroscience*, 31(Cm), pp. 195–218. doi:10.1146/annurev.neuro.31.060407.125547.

Lin, T.-Y. *et al.* (2017) 'Focal loss for dense object detection', in *Proceedings of the IEEE international conference on computer vision*, pp. 2980–2988.

Mathis, A. *et al.* (2018) 'DeepLabCut: markerless pose estimation of user-defined body parts with deep learning', *Nature Neuroscience*. Springer US, 21(September). doi: 10.1038/s41593-018-0209-y.

Michaels, J. A. and Scherberger, H. (2018) 'Population coding of grasp and laterality-related information in the macaque fronto-parietal network', *Scientific Reports*, 8(1), pp. 1–15. doi: 10.1038/s41598-018-20051-7.

Ren, S. *et al.* (2015) 'Faster r-cnn: Towards real-time object detection with region proposal networks', in *Advances in neural information processing systems*, pp. 91–99.

Schaffelhofer, S. and Scherberger, H. (2012) 'A new method of accurate hand- and arm-tracking for small primates', *Journal of Neural Engineering*, 9(2). doi: 10.1088/1741-2560/9/2/026025.

Schaffelhofer, S. and Scherberger, H. (2016) 'Object vision to hand action in macaque parietal , premotor , and motor cortices', *eLife*, 5, pp. 1–24. doi: 10.7554/eLife.15278.

Schieber, M. H. and Hibbard, L. S. (1993) 'How somatotopic is the motor cortex hand area?', *Science*, 261(5120), pp. 489–492. doi: 10.1126/science.8332915.

Sheshadri, S. *et al.* (2019) '3D reconstruction toolbox for behavior tracked with multiple cameras', 5, pp. 10–13. doi: 10.21105/joss.01849.

Simon, T. *et al.* (2017) 'Hand keypoint detection in single images using multiview bootstrapping', *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017-Janua, pp. 4645–4653. doi: 10.1109/CVPR.2017.494.

Sundaram, S. *et al.* (2019) 'Learning the signatures of the human grasp using a scalable tactile glove', *Nature*. Springer US, 569(7758), pp. 698–702. doi: 10.1038/s41586-019-1234-z.

Vargas-Irwin, C. E. *et al.* (2010) 'Decoding complete reach and grasp actions from local primary motor cortex populations.', *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 30(29), pp. 9659–9669. doi: 10.1523/JNEUROSCI.5443-09.2010.

2.3 3D reconstruction toolbox for behavior tracked with multiple cameras

Swathi Sheshadri[1,2], Benjamin Dann[1], Timo Hueser[1,3], Hansjörg Scherberger[1,2†]

[1]German Primate Center, Kellnerweg 4, 37077 Göttingen, Germany
[2]Faculty of Biology and Psychology, University of Göttingen, 37073 Göttingen, Germany
[3]Faculty of Physics, University of Göttingen, 37073 Göttingen, Germany

[†]Corresponding author. Email: hscherberger@dpz.eu

Published in JOSS: DOI: 10.21105/joss.01849

Author contributions: S.S., B.D., T.H., and H.S. designed and planned the project. S.S., programmed and documented the toolbox. T.H., B.D., and S.S performed the experiment and built the experimental setup. S.S. wrote the manuscript. All authors revised the manuscript.

The Journal of Open Source Software

3D reconstruction toolbox for behavior tracked with multiple cameras

Swathi Sheshadri[1, 2], **Benjamin Dann**[1], **Timo Hueser**[1], **and Hansjoerg Scherberger**[1, 2]

1 German Primate Center, Goettingen, Germany **2** Department of Biology and Psychology, University of Goettingen, Germany

DOI: 10.21105/ joss.01849

Software

- Review
- Repository
- Archive

Editor: Christopher R. Madan

Reviewers:

- @danasolav
- @sreschechtko

Submitted: 14 August 2019
Published: 07 January 2020

License
Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License (CC-BY).

Summary

Markerless tracking is a crucial experimental requirement for behavioral studies conducted in many species in different environments. A recently developed toolbox called DeepLabCut (DLC) (Mathis et al. (2018)) leverages Artificial Neural Network (ANN) based computer vision to make precise markerless tracking possible for scientific experiments. DLC uses a deep convolutional neural network, ResNet (He, Zhang, Ren, & Sun (2016)) pre-trained on ImageNet database (Deng et al. (2009)) and adapts it to make it applicable for behavioral tracking tasks. To track complex behaviors such as grasping with object interaction in 3D, experimental setups with multiple cameras are required. Such systems can largely benefit from a robust and easy to use camera calibration and 3D reconstruction toolbox. The current version of DLC allows 3D reconstruction of features tracked in 2D from pairs of cameras only (Nath, Mathis, Chen, Bethge, & Mathis (2019)) and is not sufficient when behavior is tracked with more than 2 cameras. Furthermore, for noisy 2D tracking conditions, such as low light or low resolution, the accuracy of tracked 3D coordinates can be improved by evaluating data from more than two cameras.

To facilitate 3D tracking of complex behaviors requiring multiple cameras (n>=2), we developed pose3d: a Matlab (The MathWorks Inc., Natick, Massachusetts) implementation for 3D reconstruction. pose3d can be divided into two main parts: camera calibration with pairwise estimation of relative camera positions to each other and 3D reconstruction with triangulation performed over 2D feature coordinates tracked from all cameras. Camera calibration refers to the estimation of intrinsic and extrinsic camera parameters including the level of magnification captured by the focal length, the distortion in the camera image caused by the lens and the location of camera in 3D coordinates. In stereo camera calibration, one of the cameras is fixed as the primary camera and the position of the secondary camera with respect to the primary camera is estimated and given as an additional parameter. Stereo camera calibration in pose3d is performed by using checkerboard presented at varied angles to the pair of cameras being calibrated to establish the correspondence between 2D image coordinates and 3D coordinates. Following calibration, the 2D feature tracked across cameras is transformed into 3D coordinates using triangulation. pose3d implements triangulation by minimizing the distance of the estimated 3D feature coordinate from the lines passing through the focal center and the feature tracked in 2D on the image plane of the cameras simultaneously.

pose3d usage and features

To use pose3d for stereo camera calibration and 3D reconstruction the users need only to edit a configuration file to enter their experiment specific details. The configuration file is extensively

The Journal of Open Source Software

commented to help users through the process of editing it. Running the main function of the repository 'main_pose3d.m' with the user's configuration file, automatically creates the folder structure required for stereo camera calibration and 3D reconstruction. Then, pose3d launches Matlab's stereoCameraCalibrator GUI sequentially for every pair of primary and secondary cameras to estimate the parameters of the cameras in the user's experimental setup. The GUI first parses through checkerboard frames recorded simultaneously from the two cameras being calibrated to select frames in which checkerboard can be detected in both cameras. At this point, the user can select the number of distortion coefficients which determines the degree of polynomial used to estimate distortion. If the lens used in the experiment causes higher distortion of images, 3 distortion coefficients can be used instead of the 2 (default) by simply selecting the radio-button on the GUI corresponding to it. After this, clicking the calibrate button on the GUI, estimates camera parameters and relative positioning of secondary camera with respect to the primary. The GUI also displays reprojection errors across calibration frames allowing users to iteratively recalibrate after removing outliers (this can be done by right-clicking on the data browser of the GUI and selecting the option remove and recalibrate). After calibration is complete, the session can be saved (by clicking the save as button). This procedure is then repeated for all camera pairs to be calibrated. pose3d prompts users throughout this procedure, providing messages on the next steps to be carried out.

After calibrating all cameras in the setup with respect to the primary camera, the 2D feature of interest tracked across cameras are corrected for distortion and the 3D coordinates estimated by triangulation. The function for triangulation across 'n' cameras is called 'triangulate_ncams'. This function is available for users in three modes that can be selected based on prior knowledge of the experimental setup. The first mode of triangulation is referred to as 'all' and uses data from all 'n' cameras that are used to track the feature in 2D. The second mode is referred to as 'avg' and performs triangulation over all pairs of cameras in the setup and provides the average over all pairs as the result. The third mode is referred to as 'best-pair' and selects the best camera pair for every time point and feature of interest for triangulation. While the first and second modes can be used for 3D reconstruction of features tracked with any software, the third mode is applicable only when tracked 2D features are associated with likelihood values, e.g., as provided by DLC.

Overall, pose3d offers a semi-automated 3D reconstruction workflow going beyond Matlab's existing functions (details in section titled 'Comparison of pose3d to existing Matlab functions') that takes users through the entire process of camera calibration, undistortion, triangulation, and post processing steps such as filtering to reduce outliers. pose3d also allows users to try different pre- and post-processing parameters and visualizes its effect on the accuracy of 3D reconstruction to help perform manual parameter tuning before saving final results. Other implementations of similar functionality as pose3d exist and use OpenCV (("Open Source Computer Vision Library," 2015)) in python: anipose (Karashchuk (2019)). However, in comparison to OpenCV based implementations, pose3d provides a user-friendly application by integrating with a feature-rich graphical user interface (GUI) for stereo camera calibration in Matlab. A more detailed comparison between pose3d and anipose is included in a subsequent section 'Comparison of pose3d and anipose'. Furthermore, we provide two demo datasets (see next paragraph) illustrating the capabilities of pose3d and a quick-view of the included features. Given the popularity of Matlab in academia and its well documented and easy to use core functions for camera calibration, we believe this toolbox will help make 3D reconstruction of tracked 2D behavior easier to use.

Demo datasets and error measurement in 3D reconstruction

Demo datasets provided in pose3d were acquired by moving a Rubik's cube on a table along different directions and by recording it simultaneously from a 5-camera tracking system. For

JOSS
The Journal of Open Source Software

the demos, we tracked the corners of the cube using DLC in the first example and manually in the second example. Manual annotations are used to mimic the usage of pose3d for any other 2D tracking software.

The key difference between the two examples is as follows. DLC, in addition to 2D tracking provides users with a likelihood value for every tracked feature that informs the users on how confident the network is about the inferred location of a particular feature of interest at any given time point. pose3d makes use of this information by applying a threshold and automatically selecting the cameras that cross this threshold for 3D reconstruction. From the 2D tracked corners we use pose3d to track corners in 3D for 1000 example frames with DLC and 20 example frames with manual annotations. Following this, we reconstruct the edges of the cube and compare it to the standard edge length of a Rubik's cube (57 mm). In the demo data using DLC based 2D annotations, we obtain on average an error of 1.39 mm in 3D reconstructed edge lengths computed over all 12 edges of the cube across 1000 example frames. For the demo data using manual annotations across 20 frames we obtain an average error of 1.16 mm over all 12 edges computed over 20 manually annotated frames. Furthermore, using 'all' mode of triangulation provided significantly better results in both of our demo datasets than the other two modes of triangulation (comparison tests for the 3 modes of 3D reconstruction included in the demo functions for reference).

Comparison of pose3d and existing Matlab functions

pose3d is a 3D reconstruction workflow integrated closely with functions from Matlab's computer vision toolbox. This workflow fixes one of the cameras in the experimental setup as primary which is crucial to ensure that the data points are within the same coordinate system across camera pairs. The stereo camera calibration part of our implementation can be considered a wrapper around functions from Matlab's computer vision toolbox, however, the triangulation function included with pose3d extends the existing triangulation function in Matlab that operates only for 2 cameras.

Comparison of pose3d and anipose

Functionally, pose3d and anipose provide users with pre- and post-processing tools for 3D reconstruction including the process of camera calibration. The main difference is that pose3d provides a user-friendly GUI utility which is missing in anipose. With pose3d, by entering the experimental details in a configuration file and running a single Matlab function ('main_pose3d.m'), the user can perform the entire procedure of 3D reconstruction. Furthermore, the integration of pose3d with the stereoCameraCalibrator GUI in Matlab makes the task of camera calibration very intuitive. This allows users to estimate camera and lens distortion parameters, select outlier frames from calibration with the click of a button. On the other hand, anipose uses OpenCV functions that can be quite difficult to use for non-experts. For instance, with pose3d, due to its integration with Matlab's stereoCameraCalibrator GUI, all pairs of images used for camera calibration are visualized along with their reprojection errors, which allows easy identification and removal of images with larger error values. With OpenCV functions, users get warning/error messages that are often quite technical in nature and can take longer for non-experts to perform root-cause analysis. In addition, pose3d provides demo datasets including example calibration images along with code to run 3D reconstructions on demo datasets as well as video tutorials to illustrate its operation. Lastly, error calculation routine is included in pose3d to quantify deviation of the reconstructed data from ground truth which is not provided in the current version of anipose. For example, in the demo dataset using Rubik's cube the 12 edges of the cubes were physically measured and compared against the reconstructed edge lengths obtained after running pose3d. Similarly, fixed lengths in the

The Journal of Open Source Software

user's experimental setups that are tracked can be compared against the measured values using the error calculation routine of pose3d.

Some other minor similarities and differences between anipose and pose3d include the following. First, both anipose and pose3d allow users to have the same set of calibration files across sessions or have a separate set per session. This can be decided by the user depending on whether the user wants to have the same calibration files across sessions or not. This is possible in pose3d by having the same experiment path (assigned to variable exp_path) and changing only the experiment name (assigned to variable exp_name) in the configuration file in case the same calibration files are used across sessions. With anipose this is implemented by having a hierarchy for the selection of calibration videos placed in different folders. Second, the triangulation approach used is the same in the two packages. However, with pose3d we have included 3 modes of 3D reconstruction including 'all', 'avg', and 'best-pair'. Of these three modes 'all' corresponds to the one implemented in anipose. With our demo dataset, we have shown that the 'all' mode gives the lowest on average error and is therefore recommended. Third, anipose allows for camera calibration using charuco boards, checker boards or aruco boards of which it recommends the usage of charuco and checker boards. Since pose3d integrates with the camera calibration GUI from MATLAB, only checker board is used for calibration.

For further technical reading on details of triangulation for 3D reconstruction please refer to our supporting document.

References

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In *CVPR09*. doi:10.1109/CVPR.2009.5206848

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778). doi:10.1109/CVPR.2016.90

Karashchuk, P. (2019). Anipose. *GitHub repository*. https://github.com/lambdaloop/anipose; GitHub. doi:10.5281/zenodo.3364758

Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., & Bethge, M. (2018). DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience*, *21*, 1281–1289. doi:10.1038/s41593-018-0209-y

Nath, T., Mathis, A., Chen, A. C., Bethge, M., & Mathis, M. W. (2019). Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nature Protocols*. doi:10.1038/s41596-019-0176-0

Open Source Computer Vision Library. (2015). https://opencv.org.

Solving for 3D coordinates χ given n sets of 2D images coordinates

Here we describe the 3D reconstruction approach implemented in our toolbox in detail. For an example point in world coordinates $[X, Y, Z]$ recorded from 'n' cameras at a given time, our implementation follows the textbook (@Hartley2003) prescribed approach to 3D reconstruction. Given we have 'n' cameras we get 'n' sets of measurements of the point we are interested in reconstructing and is given by the below system of equations for each camera 'i'.

$$\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = P_i \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{1}$$

Here the projection matrix P_i, is of the form in equation (2). It is composed of intrinsic camera parameters namely focal length and principal point and extrinsic parameters namely rotation and translation vectors. These matrices are obtained from the camera calibration process.

$$P = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \tag{2}$$

We can combine the 'n' sets of linear equations into one of the form in (3) using cross product and solve for χ

$$A\chi = 0 \ ; \ \chi = \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{3}$$

Since $[x_i, y_i, 1]$, is in the column space of P_i, cross product between vectors $[x_i, y_i, 1]$ and $[p_i^{1T}\chi, p_i^{2T}\chi, p_i^{3T}\chi]$ results in two independent equations shown below.

$$\begin{aligned} x_i \left(p_i^{3T}\chi\right) - \left(p_i^{1T}\chi\right) = 0 \\ y_i \left(p_i^{3T}\chi\right) - \left(p_i^{2T}\chi\right) = 0 \end{aligned} \tag{4}$$

Obtaining similar equations from all cameras and concatenating them along the rows of A results in a system of equations represented in equation (3) with

$$A = \begin{bmatrix} x_1 p_1^{3T} - p_1^{1T} \\ y_1 p_1^{3T} - p_1^{2T} \\ \cdot \\ \cdot \\ x_n p_n^{3T} - p_n^{1T} \\ y_n p_n^{3T} - p_n^{2T} \end{bmatrix} \tag{5}$$

Due to low image resolution and errors in camera calibration among others, there is no exact solution. Approximate solution is obtained by minimizing (3) in the least squared sense. To avoid trivial solution, constrained equation with Lagrange multiplier is constructed as in equation (6) and minimized (@Inkila2005).

$$L(x, \lambda) = |A\chi|^2 + \lambda \left(1 - |\chi|^2\right) \tag{6}$$

Taking partial derivative with respect to χ and setting it to 0 to minimize L, we get

$$A^T A\chi = \lambda\chi \tag{7}$$

Substituting (7) in equation (6), we get

$$L = |A\chi|^2 + \lambda \left(1 - |\chi|^2\right) = \chi^T A^T A\chi + \lambda \left(1 - \chi^T\chi\right) \tag{8}$$

$$L = \chi^T \lambda\chi + \lambda - \lambda\chi^T\chi = \lambda \tag{9}$$

From equation (9), we see that the χ that minimizes (3) in the least squared sense is the Eigen vector of $A^T A$ associated with its lowest Eigen value.

References

[1] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[2] Keijo Inkilä. Homogeneous least squares problem. *Photogrammetric Journal of Finland*, 19(2):34–42, 2005.

README.md

3D reconstrution toolbox for behavior tracked with multiple cameras

What is pose3d?

pose3d is implemented in Matlab (The MathWorks Inc., Natick, Massachusetts) for 3D reconstruction of features tracked in 2D using DeepLabCut (DLC) or any other tracking software. It provides a semi-automated 3D reconstruction workflow that takes users through the process of camera calibration, undistortion, triangulation as well as post processing steps such as filtering to reduce outliers. In addition to providing an easy to use workflow, the key component in our implementation is the 'n' camera triangulation function that allows users to select 2D tracked data from the best pair of cameras for every feature and time point using an automated selection criterion or use data from all cameras for 3D reconstruction.

Table of Contents

- Why pose3d?
- Running demos
- Using pose3d for your data
- Data requirements for pose3d
- Checkerboard images or videos for calibration
- Make your own movie with 3D reconstruction
- Dependencies
- Video tutorial and troubleshooting
- Installation
- Citation
- Acknowledgements

Why pose3d?

Markerless tracking is a crucial experimental requirement for behavioral studies conducted in many species in different enviroments. A recently developed toolbox called DeepLabCut (DLC) leverages Artificial Neural Network (ANN) based computer vision to make precise markerless tracking possible for scientific experiments. To track complex behaviors such as grasping with object interaction in 3D as illustrated in Figure 1, experimental setups with multiple cameras have to be developed. Development of such systems can largely benefit from a robust and easy to use camera calibration and 3D reconstruction toolbox. To this end, we developed pose3d, a semi-automated 3D reconstruction toolbox in Matlab. Given the popularity of Matlab in academia, we believe this toolbox will help make 3D reconstruction of 2D tracked behavior easier for researchers to use.

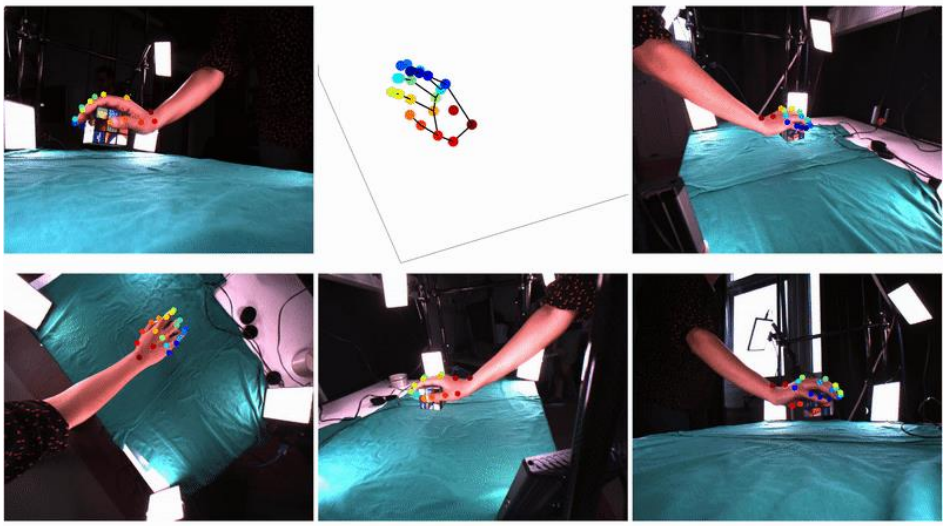

Figure 1: Reconstruction in 3D of grasping behavior tracked in 2D using DeepLabCut

Running demos

1. Download or clone our pose3d repository. From the main folder of the repository run ./Codes/demo_DLC2d.m to perform 3D reconstruction of the corners of a Rubik's cube tracked in 2D using DLC.
   -- Since undistortion, camera calibration and 3D reconstruction have already been done and saved for the demo, click 'no' when message boxes pop-up and ask if you want to save. 2D tracked data and videodata used for this demo are included in the DemoData folder for reference.
2. Similiar to the first demo from the main folder of the repository, run ./Codes/demo_other2d.m to perform 3D reconstruction of the corners of a Rubik's cube tracked in 2D using any other 2D tracking software. Here we mimick other software by manual 2D tracking.

Using pose3d for your data

Follow the below steps to perform 3D reconstruction of your 2D tracked data.

1. Edit ./Codes/template_config_file.m file to enter your project specific information.
   -- pose3d allows users to reuse camera calibration files across recording sessions. This is useful when your cameras are fixed in a place and you capture behavior of interest over several recording sessions. To reuse calibration files across sessions, when editing the configuration file use the same name for exp_path (experiment path) variable changing only the exp_name (experiment name) variable. To use different calibration data, provide a new entry to the exp_path variable. In other words, every experiment within the same experiment path uses the same calibration data. Please read the comments in the file template_config_file.m for details on the other required and optional variables for pose3d while filling in your experiment details into the configuration file.
2. Run ./Codes/main_pose3d.m with your config file to save 3D reconstructed data for your project.
3. The main_pose3d script is semi-automated. When user input is required message boxes pop-up in Matlab and wait for your response. Follow the messages to perform 3D reconstruction.
   -- Upon running main_pose3d.m, you will find a folder created for all your experiments with a sub-folder for your current experiment using names provided by you in the config file. Calibration images will be extracted from calibration videos (or directly copied if you already have calibration images) and sorted for you as illustrated in ./DemoExperiments/Imagesforcalibration/ folder. Furthermore, within your current experiment folder, a folder for 3D reconstructed data called Data3d will be created where results of 3D reconstruction will be saved as Data3d.mat file. In case you set undistortion and filter options in your config file, a folder each for 2D undistorted data called UndistortedData2d and for filtered 3D data called FilteredData3d will be created.

4. After saving 3D reconstructed data, if you have chosen to plot the results by setting plotresults variable in the config file, 3D reconstructed data will be displayed adjacent to the provided movie file recorded from primary camera. Enter the path to the movie file recorded from primary camera in the config file to visualize 3D reconstructed data along side the movie recorded during the experiment. If you have recorded images and tracked behavior across images: provide the path to the folder where you have saved the image files from primary camera. The folder should exclusively contain the 2D tracked images from the primary camera (in .png, .tiff, .jpeg or any other format compatible with imread function in Matlab). Furthermore, name the images with numbers to help pose3d match the order of the images to that of the 2D tracked information provided in the .csv file. For example, if you have 10 frames, save them as frame01.png, frame02.png, ... frame10.png. Similarly, if you have 100 frames, save them as frame001.png, frame002.png, ...frame100.png.
5. If you have chosen to calculate error in 3D reconstruction by setting calc_error variable to 1, average and standard deviation in error will be displayed at the successful termination of main_pose3d. Calculation of errors, however, requires users to provide ground-truth and drawline variables in the config file (detailed comments are available in the config file to help fill this variable). If ground truth is not available for a certain line segment provide NaN for it to inform pose3d to ignore it during error calculation. Further details on reconstruction error corresponding to individual frames and line segments can also be accessed from the workspace variable errorinrecon (frames x linesegments x mode).

Data requirements for pose3d

Note: If you are using DLC for 2D tracking from 2 or more cameras and have saved the .csv files available from DLC after analyzing videos you only need step 3 from the following list.

1. Record behavior of interest from multiple cameras (n >= 2) and track features of interest across these cameras using DLC or any other 2D tracking software.
2. Save 2D tracked data in .csv format and keep these files accessible from the computer on which you want to use pose3d toolbox.
3. Record checkerboard calibration videos by fixing one of your setup cameras as primary. Primary camera should be selected such that you can obtain checker board images simultaneously with the primary and every other camera in full view. For instance, if you have 3 cameras, let us call one as primary and the remaining cameras as secondary 1 and secondary 2. From this setup, first record checkerboard videos with cameras primary and secondary 1 simultaneously. Next record checkerboard videos with cameras primary and secondary 2 simultaneously. Save these videos and have them accessible from your computer.
4. If you are using software other than DLC follow the below guidelines to prepare your 2D tracked data for pose3d.
   Pose3d requires one .csv file per camera, each containing the 2D coordinates of the tracked features of interest from that camera. Please maintain the order of the features the same in all the .csv files you prepare for all the cameras in your setup.
   Case 1: When you have 2D coordinates of the tracked features and do not have likelihood values per feature and time point as given by DLC. In this case, set the usingdlc variable in your configuration file to 0 and when preparing the .csv files, include the x and y coordinates of the features you have tracked along the columns of your .csv file. For example, if you have 2 features. Then first column contains x coordinate of feature 1, second column contains y coordinate of feature 1 followed by the x and y coordinates of feature 2 along columns 3 and 4. The rows of your .csv file correspond to frames of the video or time points, with the first row containing the tracked 2D coordinates of the tracked features from time-point or frame 1 to the nth row containing the 2D coordinates of the tracked features from time-point or frame n. Optionally, you can include row and column headings as text elements for your reference in the .csv files, however pose3d does not make use of it.
   Case 2: When you have 2D coordinates of the tracked features and also have likelihood values associated with each feature of interest at every time point. In this case the data you have resembles the 2D tracked data in the case of DLC users. So, set the usingdlc variable in the configuration file to 1 and build your .csv file as described below. Fill the first column of your .csv file with serial numbers going from 0 to n-1 (where n corresponds to the number of frames). In subsequent columns include the x and y coordinates and the likelihood values of the features you have tracked. For example, if you have 2 features and 10 frames or time points. Then first column holds numbers going from 0 to 9; second column contains x coordinates of feature 1; third column contains y coordinates of feature 1; followed by the fourth column which holds its likelihood values. The likelihood values can range from 0 to 1. Following this fifth, sixth and seventh columns contain, x and y coordinates and the likelihood values of feature 2 respectively. Optionally, here as well, you can include row and column headings as text elements for your reference in the .csv files, however pose3d does not make use of it.

Checkerboard images or videos for calibration

1. Print checkerboard for camera calibration -- This can be done by first typing 'open checkerboardPattern.pdf' in Matlab command line and then printing out the checkerboard layout. Please make sure the layout is attached to a smooth and stable surface before using it for calibration.
2. Present the checkboard to your primary and secondary cameras simultaneously so that the whole frame of the checkerboard is within the field of view of the two cameras. Acquire images or videos of the checkboard being presented at varied angles and distances from the camera.
3. Matlab's stereoCameraCalibrator GUI recommends using between 10 and 20 frames for calibration. However, we recommend collecting between ~50-100 frames (especially in low light conditions where the checkerboard would be harder to detect), since some of the frames automatically get rejected by the GUI based on the inability to detect whole checkboard or due to the tilt angle of the board being > 45 degree relative to the camera plane. Further tips on acquiring calibration files can be obtained at: Mathworks' Camera Calibration

Make your own movie with 3D reconstruction

The default plotting function in pose3d is designed for quick visualization of results alongside annotated 2D images from primary camera. The default plotter has the option to visualize every nth frame based on the nskip variable in the config file and also prompts after every minute to decide if you want to continue watching the results movie or are happy with the results. In case of high data volume exceeding working memory available for Matlab, due to large number of cameras, high resolution images, higher frame rate etc., the plotter defaults to displaying only the 3D reconstructed data in a frame by frame format. Alternatively, when you have high data volume or want to visualize 2D videos from multiple cameras, helper function called **make_illustrative_movie.m** can be used. This function allows you to flexibly select any number of perspectives you want to visualize alongside 3D reconstructed results. make_illustrative_movie.m is a standalone function and can be used by editing it to include the config file that was used to do 3D reconstruction and by providing paths to the videos and .csv files holding 2D tracked videos for the respective videos of your experiment. Successful execution of make_illustrative_movie.m function, plots all frames from the selected videos, alongside 3D reconstructed results and saves the results in .avi format in a subfolder called Videos in your experiment folder.

Dependencies

In-built functions of Matlab, Computer Vision Toolbox. Code has been tested on Matlab 2018b across linux, MAC and Windows operating systems.

Video tutorial and troubleshooting

Video tutorials to use pose3d can be found at the below links:
Case 1: 3D reconstruction when calibrating with images
Case 2: For calibration with recorded videos
Pose3d incorporates message-based interaction to inform the user when problems are detected while using pose3d. This includes checking entries made by the user in the configuration file to check for congruency in the entered values. For instance, if the user sets number of cameras in the setup to 5 but provides lesser number of .csv files containing 2D tracked data, a message informing the user about this problem is provided. However, some problems can occur without producing explicit error messages and are listed below to help users with troubleshooting.

1. Mismatch in the correspondence between 2D tracked .csv files and calibration files.
   -- This issue can occur when the user is editing the configuration file to enter his/her experiment details. When entering file names into variables secondary2D_datafullpath and calibvideos_secondary, the order in which the files corresponding to secondary cameras is entered must be maintained. Otherwise, there will be failure in 3D reconstruction since data from one camera will be transformed using camera matrix of a different camera. To avoid this problem, ensure that the order of 2D data files and calibration files is maintained. See the comments in the template config file for more details.
2. Poor estimation of the relative positioning of the camera and lens properties such as level of magnification and image distortion.
   -- This can happen if calibration frames are too few and the workspace is not covered sufficiently (at varied distances and angles)

by the frames used for calibration. To avoid this case, present the checker board in different angles and record it in full view from both the primary and secondary cameras by moving it about the entire workspace. If you are making calibration videos, ensure that the whole checkboard is clearly visible in most if not all frames. In addition, check the re-projection errors in the stereo camera calibration GUI to remove outliers (right click on the images displayed on the data browser of the GUI and select remove and recalibrate) for better results. If you are unsure of the kind of calibration videos or images to take, please see our example calibration images in the ./DemoExperiments/Imagesforcalibration/ folder. See section Checkerboard images or videos for calibration in this readme for more details on calibration.

3. Camera movement between recording calibration videos and behavior.
   -- Ensure that your cameras are fixed between recording sessions. Record and use new calibration videos everytime you suspect that the cameras in your setup have moved.
4. Positioning of the cameras is not optimal or number of cameras is not sufficient for the behavior of interest.
   -- For every feature of interest you want to track, you must be able to perform 2D tracking reliably from at least two cameras. Ensure to have sufficent number of cameras to track the behavior of interest and position the cameras such that they cover the tracking workspace as much as possible.
5. Our toolbox is designed to work with stereoCameraCalibrator GUI in Matlab. The GUI requires simulataneously acquired images from the two cameras to have the same name and saved it different folders. In case you have acquired calibration videos, all you need to do is edit the config file to enter the path and name of the calibration videos and pose3d automatically does the rest for you. In case you have acquired images from calibration, please have your images labelled as shown in ./DemoExperiments/Imagesforcalibration/ folder. pose3d automatically goes through every primary and secondary camera pairs and prompts you to select all calibration images per camera.

Installation

To install this toolbox, add all contents of this repository to Matlab path.

Citation

Sheshadri et al., (2020). 3D reconstruction toolbox for behavior tracked with multiple cameras. Journal of Open Source Software, 5(45), 1849, https://doi.org/10.21105/joss.01849
For questions on the toolbox please contact us at swathishesh@gmail.com or HScherberger@dpz.eu

Acknowledgements

We are very thankful to the JOSS editor Christopher Madan, and reviewers Dana Solav and Sasha Reschechtko whose feedback helped us improve pose3d immensely for public use. We are also very thankful to Mackenzie Mathis and Alexander Mathis for their help in getting us going with vision based tracking using DLC. Discussions with several colleagues at the German Primate Center including Andrej Fillipow, Michael Berger, Sebastian Mueller, Attila Trunk and Daniela Buchwald were very useful for the development of our tracking experimental setup. Thanks also to Viktorija Schek for testing out some of our implementations. Our implementation in Matlab was inspired partly by Anipose - a python toolbox for 3D reconstruction.

3 General Discussion

3.1 Summary

In the first part (Chapter 2.1) of this thesis, neuron-resolved dynamics of oscillatory network structure was examined over time and across different conditions. To this end, spikes and LFPs were recorded simultaneously from macaques performing a delayed grasping task. The task included different context-, grip-types as well as different behavioral epochs, which enabled the investigation of dynamic changes of the oscillatory network structure during flexible sensorimotor transformations. To make the comparison of network structure across conditions with different firing rates possible, pairwise phase consistency, a rate unbiased measure of spike-field phase locking was used. Separate groups of neurons oscillatory synchronized in the beta and low frequencies resulting in separate subnetworks. These frequency specific subnetworks were active during different behavioral epochs, suggesting the frequencies to be mutually exclusive at the level of neurons. Importantly, both frequency subnetworks reconfigured for different task conditions, which was captured predominantly at the level of single neurons. Together, these findings suggest that the oscillatory network structure might provide a coordinative framework for flexible sensory-motor transformations.

In Chapter 2.2 of this thesis, the developmental and implementational details of a markerless paradigm to track unconstrained grasping behavior in primates were described. Although there exist other hand tracking solutions (Zhang *et al.*, 2020), they did not include object interactions as they focused on tracking hand gestures for applications such as sign language decoding. To understand cortical control of grasping movements, it is crucial to track hand kinematics during hand-object interactions. A recent development allows precise markerless hand tracking during object interactions (Simon *et al.*, 2017). However, it requires over 30 cameras mounted in a specially designed pan-optic studio. The solution proposed in this thesis enables tracking 22 keypoints of the hand from behavioral videos acquired

simultaneously from as few as 5 cameras. Video frames that had a high signal to noise ratio were acquired by sufficiently and uniformly lighting the experimental setup. Frame rate, resolution, and exposure settings of the camera, as well as lens properties such as magnification and distortion, were chosen to ensure unblurred images with resolvable hand keypoints. In addition, a technique to estimate keypoints in occluded views by reprojecting 3D reconstructed keypoints from unoccluded views was developed. This improved the quality of manual annotations, which was then used to train a convolutional neural network in DeepLabCut (Mathis *et al.*, 2018). The network training was repeated on successively larger training datasets, which were generated automatically by selecting frames annotated by the network trained in the previous iteration. A reprojection error based criterion was used to ensure the quality of the network annotated frames that were selected for training. Using this technique, we demonstrated precise hand tracking during an unconstrained human grasping task, which involved a wide variety of object interactions. A pilot study was also carried out in macaques to demonstrate the adaptability of the markerless hand tracking system to this animal model. The results of these validation experiments clearly demonstrated the technical capabilities of the proposed solution. Notably, the ability to accurately capture unconstrained hand kinematics opens up the possibility to carry out a wider range of more natural lab experiments.

In Chapter 2.3, a Matlab toolbox called pose3d was described. To date, DLC allows 3D reconstruction of feature coordinates tracked in 2D from pairs of cameras only (Nath *et al.*, 2019). Therefore, it is not sufficient when behavior is tracked with multiple (n>2) cameras. Multiple cameras are necessary to track keypoints in 3-dimensional coordinates, particularly during complex behaviors involving high degrees of freedom such as the grasping movements of the hand. Pose3d was developed to address this gap and provides a user-friendly, semi-automated graphical user interface that facilitates camera calibration, distortion reduction, and triangulation of keypoints tracked in 2D. Furthermore, some post-processing filtering and annotated movie generation tools are included in pose3d. This toolbox integrates

well with DLC and will help make 3D reconstruction of tracked behavior easier to implement.

3.2 Outlook

Complex behaviors require information to be coordinated within and between brain areas. A global coordinative role has been thought to be played by oscillatory synchrony, facilitating selective and flexible routing of information in the brain (Buzsaki, 2006; Fries, 2015). However, the exact mechanism by which oscillations achieve flexible information processing remains highly debated. Some insights have been gained from gamma oscillations observed in the visual cortex in which selective modulation of attention has been linked to gamma synchronization (Fries, 2001). In this context, the theory of communication through coherence (CTC) proposed effective information transmission between pre- and post-synaptic neuronal groups that rhythmically synchronized with an appropriate phase-relationship (Figure 13A). Guided by the theory of CTC, selective communication between brain areas was examined by studying coherence between field potentials (Bosman *et al.*, 2012). However, CTC has recently been questioned by researchers who argued that the LFPs include local as well as afferent synaptic potentials (Pesaran *et al.*, 2018; Schneider *et al.*, 2020) making field-field coherence difficult to interpret. That is, when two brain areas are anatomically connected, the sending area will, by default, be coherent with the part of the LFP signal in the receiving area to which it contributes. Thereby, any mechanism such as attention that increases communication between two areas would also increase coherence as a byproduct. This synaptic mixing effect was demonstrated to result in narrow peaks in the coherence spectrum, despite no true oscillatory coupling (Schneider *et al.*, 2020). Furthermore, CTC requires a mechanism to generate reliable phase synchronization between neuronal signals, which has so far remained elusive for neocortical rhythms (Schneider *et al.*, 2020).

Alternatively, a relatively simpler proposal was made suggesting oscillations

as a selection mechanism to enable different subnetworks of neurons in a task-dependent manner (Canolty, Ganguly and Carmena, 2012; Womelsdorf, Westendorff and Ardid, 2013). Canolty and others found a reliable mapping referred to as cross-level coupling between beta oscillations and firing rate of neurons (Figure 13B).

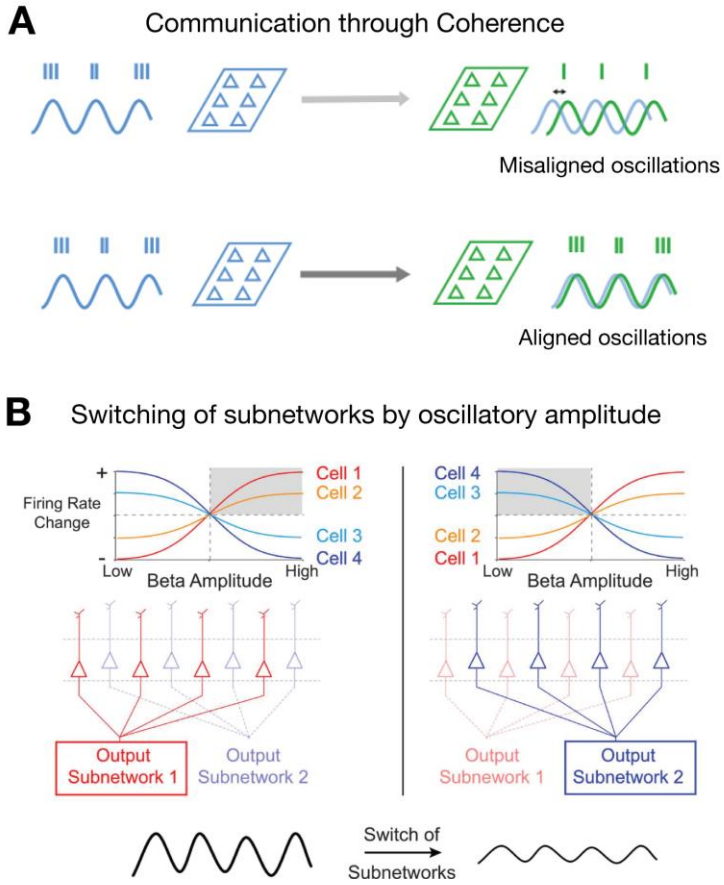

Figure 13: Comparing two schemes of selective and flexible information processing. A. Communication through coherence scheme highlighting the aligned case (bottom), where the pre- (blue) and post-synaptic (green) populations have an appropriate phase relationship. In such a case the efficacy of communication between the two populations is enhanced. The misaligned case, which weakens communication is also shown (top). **B.** Cross-level coupling scheme in which the firing rate of neurons is linked to the amplitude of beta band LFP oscillations. When the amplitude of the beta oscillations is higher (left), cells 1 and 2 have higher firing rates, when the amplitude of beta oscillations is lower, cells 3 and 4 have higher firing rate. A adapted from Kohn *et al.*, (2020) and B from Womelsdorf *et al.*, (2013)

The strength of oscillatory synchrony at the area-level was directly related to the firing rate of individual neurons in the same area. Interestingly, the firing rate of one group of neurons was higher during phases of strong oscillatory synchrony, and the

firing rate of another group of neurons was higher during phases of weak oscillatory synchrony. While different coordinative schemes have been proposed, it is noteworthy that they are not mutually exclusive (Kohn *et al.*, 2020). The above-discussed scheme for the selection of subnetworks by cross-level coupling is compatible with the theory of cell assemblies (Buzsáki, 2010). Importantly, modelling studies that simulate networks of inter-connected neurons are required to compare the different coordinative schemes and verify their biological plausibility. However, empirical evidence on the dynamics of oscillatory network structure to support such modelling studies is still largely lacking at the level of single neurons.

Filling this gap, in Chapter 2.1 of this thesis, subnetworks of neurons oscillating in specific frequencies during different behavioral states were identified. Furthermore, the identified subnetworks reconfigured for different types of sensorimotor transformations. These findings suggest oscillatory synchrony not only as a mechanism for the selection of neurons but also as a flexible routing mechanism for context-dependent behavior.

One major criticism against oscillations is that it reduces the information capacity of the system (Shadlen and Newsome, 1998). While this was perhaps a valid criticism in the past, emerging evidence has demonstrated low dimensional task-relevant manifolds in the brain (Churchland *et al.*, 2010; Kaufman *et al.*, 2014; Gallego *et al.*, 2017). However, it can be argued that the tasks that have identified low dimensional manifolds were rather simple, thereby explaining the observed low dimensional neural control (Gao *et al.*, 2017). To rule out the possibility of the observed low dimensionality merely reflecting task simplicity, the range of studied behaviors has to be expanded.

In this thesis (Chapter 2.1), to examine the role of oscillatory synchrony during sensorimotor transformations, a delayed grasping task with two context- and two grip-type variants was studied. This represents only a small subset of all the grasps that the primate uses during its everyday interactions with the environment. One potential way to expand the range of studied grasps would be to study free behaviors such as unconstrained object interactions and self-feeding in primates. If oscillatory

synchrony is indeed facilitating the coordination of visuomotor transformations, then the results presented in this thesis must continue to be valid in a task that explores a larger variety of hand movements.

An important technical requirement to study unconstrained grasping movements and object interactions is the precise tracking of hand kinematics. Sensors built into the grasping setup were sufficient to determine the executed grip type during the delayed grasping task studied here. However, to track free grasping movements and object interactions, a more advanced tracking solution is required. The tracked hand kinematics would then allow the identification of elementary grasp components during free grasping and object interaction activity. Furthermore, the relationship between the wide range of free grasping movements and the simultaneously acquired neuronal signals can be studied. However, the tracking solution must meet a number of requirements to be applicable for such a task. First, hand kinematics must be tracked reliably during the experiment. Second, the acquisition of hand kinematics must not interfere with neural recordings. Third, hand kinematics must allow the acquisition of unconstrained natural behaviors. Although a kinematic data glove (Schaffelhofer and Scherberger, 2012) provides reliable hand tracking, it restricts the range of hand movements. Furthermore, there is a very high chance for the sensors to get destroyed during object interactions or self-feeding movements, which are both parts of unconstrained grasping movements. The markerless tracking solution developed as a part of this thesis (Chapters 2.2, 2.3) satisfies the requirements to enable precise tracking of unconstrained grasping behavior.

Overall, by combining the neuron-resolved network analysis with precise hand tracking during a task consisting of a larger repertoire of grasping movements, further insights about the underlying grasping network can be drawn. Empirical findings gained from such studies provide the basis of biological plausibility for future modelling studies investigating potential mechanisms of flexible information processing in the brain.

4 Bibliography

Achermann, P. and Borbély, A. A. (1997) 'Low-frequency (< 1 hz) oscillations in the human sleep electroencephalogram', *Neuroscience*, 81(1), pp. 213–222. doi: 10.1016/S0306-4522(97)00186-3.

Aggarwal, V. *et al.* (2013) 'State-based decoding of hand and finger kinematics using neuronal ensemble and LFP activity during dexterous reach-to-grasp movements', *Journal of Neurophysiology*, 109(12), pp. 3067–3081. doi: 10.1152/jn.01038.2011.

Ahrens, M. B. *et al.* (2013) 'Whole-brain functional imaging at cellular resolution using light-sheet microscopy', *Nature Methods*, 10(5), pp. 413–420. doi: 10.1038/nmeth.2434.

Allen, E. A. *et al.* (2014) 'Tracking whole-brain connectivity dynamics in the resting state', *Cerebral Cortex*, 24(3), pp. 663–676. doi: 10.1093/cercor/bhs352.

Alom, Z. *et al.* (2018) 'The History Began from AlexNet: Comprehensive Survey on Deep Learning Approaches'. Available at: https://arxiv.org/pdf/1803.01164.pdf.

Amzica, F. and Steriade, M. (1998) 'Electrophysiological correlates of sleep delta waves', *Electroencephalography and Clinical Neurophysiology*, 107(2), pp. 69–83. doi: 10.1016/S0013-4694(98)00051-0.

Aoi, M. C. *et al.* (2015) 'Rate-adjusted spike-LFP coherence comparisons from spike-train statistics', *Journal of Neuroscience Methods*. Elsevier B.V., 240, pp. 141–153. doi: 10.1016/j.jneumeth.2014.11.012.

Arce-McShane, F. I. *et al.* (2016) 'Primary motor and sensory cortical areas communicate via spatiotemporally coordinated networks at multiple frequencies', *Proceedings of the National Academy of Sciences*, 113(18), pp. 5083–5088. doi: 10.1073/pnas.1600788113.

Bassett, D. S. *et al.* (2009) 'Cognitive fitness of cost-efficient brain functional networks', *Proceedings of the National Academy of Sciences of the United States of America*, 106(28), pp. 11747–11752. doi: 10.1073/pnas.0903641106.

Bastos, A. M. *et al.* (2015) 'Visual areas exert feedforward and feedback influences through distinct frequency channels', *Neuron*, 85(2), pp. 390–401. doi: 10.1016/j.neuron.2014.12.018.

Bastos, A. M. and Schoffelen, J.-M. (2016) 'A Tutorial Review of Functional Connectivity Analysis Methods and Their Interpretational Pitfalls', *Frontiers*

in Systems Neuroscience, 9(January), pp. 1–23. doi: 10.3389/fnsys.2015.00175.

Batista, A. P. *et al.* (1999) 'Reach plans in eye-centered coordinates', *Science*, 285(July), pp. 257–260.

Baumann, M. A., Fluet, M. C. and Scherberger, H. (2009) 'Context-specific grasp movement representation in the macaque anterior intraparietal area', *Journal of Neuroscience*, 29(20), pp. 6436–6448. doi: 10.1523/JNEUROSCI.5479-08.2009.

Bengio, Y., Simard, P. and Frasconi, P. (1994) 'Learning long-term dependencies with gradient descent is difficult', *IEEE Transactions on Neural Networks*, 5(2), pp. 157–166. doi: 10.1109/72.279181.

Berger, H. (1929) 'Über das elektroenkephalogramm des menschen', *Archiv für psychiatrie und nervenkrankheiten*, 87(1), pp. 527–570.

Berger, M., Agha, N. and Gail, A. (2020) 'Wireless recording from unrestrained monkeys reveals motor goal encoding beyond immediate reach in frontoparietal cortex', *eLife*, 9, pp. 1–29. doi: 10.1101/305334.

Bock, D. D. *et al.* (2011) 'Network anatomy and in vivo physiology of visual cortical neurons', *Nature*. Nature Publishing Group, 471(7337), pp. 177–184. doi: 10.1038/nature09802.

Borra, E. *et al.* (2017) 'The macaque lateral grasping network: A neural substrate for generating purposeful hand actions', *Neuroscience and Biobehavioral Reviews*. Elsevier Ltd, 75, pp. 65–90. doi: 10.1016/j.neubiorev.2017.01.017.

Bosman, C. A. *et al.* (2012) 'Attentional Stimulus Selection through Selective Synchronization between Monkey Visual Areas', *Neuron*, 75(5), pp. 875–888. doi: 10.1016/j.neuron.2012.06.037.

Bragin, A. *et al.* (1999) 'High-frequency oscillations in human brain', *Hippocampus*, 9(2), pp. 137–142. doi: 10.1002/(SICI)1098-1063(1999)9:2%3C137::AID-HIPO5%3E3.0.CO;2-0.

Brovelli, A. *et al.* (2004) 'Beta oscillations in a large-scale sensorimotor cortical network: Directional influences revealed by Granger causality', *Proceedings of the National Academy of Sciences of the United States of America*, 101(26), pp. 9849–9854. doi: 10.1073/pnas.0308538101.

Buzsaki, G. (2006) *Rhythms of the Brain*. Oxford University Press.

Buzsáki, G. *et al.* (1992) 'High-frequency network oscillation in the hippocampus', *Science*, 256(5059), pp. 1025–1027. doi: 10.1126/science.1589772.

Buzsáki, G. (2005) 'Theta rhythm of navigation: Link between path integration and landmark navigation, episodic and semantic memory', *Hippocampus*, 15(7), pp. 827–840. doi: 10.1002/hipo.20113.

Buzsáki, G. (2010) 'Neural Syntax: Cell Assemblies, Synapsembles, and Readers', *Neuron*, 68(3), pp. 362–385. doi: 10.1016/j.neuron.2010.09.023.

Buzsáki, G., Anastassiou, C. A. and Koch, C. (2012) 'The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes', *Nature Reviews Neuroscience*. Nature Publishing Group, 13(6), pp. 407–420. doi: 10.1038/nrn3241.

Buzsáki, G. and Christen, Y. (2016) *Micro-, meso-and macro-dynamics of the brain*. Springer Nature.

Buzsáki, G. and Draguhn, A. (2004) 'Neuronal oscillations in cortical networks', *Science*, 304(5679), pp. 1926–1929. doi: 10.1126/science.1099745.

Buzsáki, G. and Wang, X.-J. (2012) 'Mechanisms of Gamma Oscillations', *Annual review of neuroscience*, 35, pp. 203–225. doi: 10.1146/annurev-neuro-062111-150444.

Canolty, R. T. *et al.* (2010) 'Oscillatory phase coupling coordinates anatomically dispersed functional cell assemblies', *Proceedings of the National Academy of Sciences*, 107(40), pp. 17356–17361. doi: 10.1073/pnas.1008306107.

Canolty, R. T., Ganguly, K. and Carmena, J. M. (2012) 'Task-Dependent Changes in Cross-Level Coupling between Single Neurons and Oscillatory Activity in Multiscale Networks', *PLoS Computational Biology*, 8(12). doi: 10.1371/journal.pcbi.1002809.

Cao, Z. *et al.* (2017) 'Realtime multi-person 2D pose estimation using part affinity fields', *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017-Janua(Xxx), pp. 1302–1310. doi: 10.1109/CVPR.2017.143.

Caton, R. (1875) 'Electrical Currents of the Brain', *The Journal of Nervous and Mental Disease*, 2(4). Available at: https://journals.lww.com/jonmd/Fulltext/1875/10000/Electrical_Currents_of_the_Brain.13.aspx.

Churchland, M. M. *et al.* (2010) 'Cortical Preparatory Activity: Representation of Movement or First Cog in a Dynamical Machine?', *Neuron*. Elsevier Inc., 68(3), pp. 387–400. doi: 10.1016/j.neuron.2010.09.015.

Cisek, P. (2007) 'Cortical mechanisms of action selection: The affordance competition hypothesis', *Philosophical Transactions of the Royal Society B: Biological*

Sciences, 362(1485), pp. 1585–1599. doi: 10.1098/rstb.2007.2054.

Cohen, M. R. and Kohn, A. (2011) 'Measuring and interpreting neuronal correlations.', *Nature neuroscience*. Nature Publishing Group, 14(7), pp. 811–819. doi: 10.1038/nn.2842.

Cole, M. W. *et al.* (2014) 'Intrinsic and task-evoked network architectures of the human brain', *Neuron*. Elsevier Inc., 83(1), pp. 238–251. doi: 10.1016/j.neuron.2014.05.014.

Dann, B. *et al.* (2016) 'Uniting functional network topology and oscillations in the fronto-parietal single unit network of behaving primates', *eLife*, 5, pp. 1–26. doi: 10.7554/eLife.15719.

Dann, B. (2017) 'Encoding , coordination , and decision making in the primate fronto-parietal grasping network'. Georg-August-Universität Göttingen.

Davis, J. and Shah, M. (1994) 'Recognizing Hand Gestures', *European Conference on Computer Vision*. Springer, Berlin, Heidelberg.

Dean, H. L., Hagan, M. A. and Pesaran, B. (2012) 'Only coherent spiking in posterior parietal cortex coordinates looking and reaching.', *Neuron*. Elsevier Inc., 73(4), pp. 829–41. doi: 10.1016/j.neuron.2011.12.035.

Deng, J. *et al.* (2009) 'Imagenet: A large-scale hierarchical image database', in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, pp. 248–255.

Deng, L. and Yu, D. (2013) 'Deep learning: Methods and applications', *Foundations and Trends in Signal Processing*, 7(3–4), pp. 197–387. doi: 10.1561/2000000039.

Dhamala, M., Rangarajan, G. and Ding, M. (2008) 'Analyzing Information Flow in Brain Networks with Nonparametric Granger Causality', *NeuroImage*, 41(2), pp. 354–362. doi: 10.1016/j.neuroimage.2008.02.020.

Dipietro, L., Sabatini, A. M. and Dario, P. (2008) 'A survey of glove-based systems and their applications', *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, 38(4), pp. 461–482. doi: 10.1109/TSMCC.2008.923862.

Engel, A. K. and Fries, P. (2010) 'Beta-band oscillations—signalling the status quo?', *Current Opinion in Neurobiology*, 20(2), pp. 156–165. doi: 10.1016/j.conb.2010.02.015.

Erol, A. *et al.* (2007) 'Vision-based hand pose estimation: A review', *Computer Vision and Image Understanding*, 108(1–2), pp. 52–73. doi:

10.1016/j.cviu.2006.10.012.

Fagg, A. H. and Arbib, M. A. (1998) 'Modeling parietal-premotor interactions in primate control of grasping', *Neural Networks*, 11(7–8), pp. 1277–1303. doi: 10.1016/S0893-6080(98)00047-1.

Fluet, M. C., Baumann, M. A. and Scherberger, H. (2010) 'Context-specific grasp movement representation in macaque ventral premotor cortex', *Journal of Neuroscience*, 30(45), pp. 15175–15184. doi: 10.1523/JNEUROSCI.3343-10.2010.

Fogassi, L. *et al.* (2001) 'Cortical mechanism for the visual guidance of hand grasping movements in the monkey: A reversible inactivation study', *Brain*, 124(3), pp. 571–586. doi: 10.1093/brain/124.3.571.

Fries, P. (2001) 'Modulation of Oscillatory Neuronal Synchronization by Selective Visual Attention', *Science*, 291(5508), pp. 1560–1563. doi: 10.1126/science.1055465.

Fries, P. (2005) 'A mechanism for cognitive dynamics: neuronal communication through neuronal coherence.', *Trends in cognitive sciences*, 9(10), pp. 474–80. doi: 10.1016/j.tics.2005.08.011.

Fries, P. (2015) 'Rhythms for Cognition: Communication through Coherence', *Neuron*. Elsevier Inc., 88(1), pp. 220–235. doi: 10.1016/j.neuron.2015.09.034.

Gallego, J. A. *et al.* (2017) 'Neural Manifolds for the Control of Movement', *Neuron*. Elsevier Inc., 94(5), pp. 978–984. doi: 10.1016/j.neuron.2017.05.025.

Gallese, V. *et al.* (1994) 'Deficit of hand preshaping after muscimol injection in monkey parietal cortex', *NeuroReport*, 5(12). Available at: https://journals.lww.com/neuroreport/Fulltext/1994/07000/Deficit_of_hand_preshaping_after_muscimol.29.aspx.

Gao, P. *et al.* (2017) 'A theory of multineuronal dimensionality, dynamics and measurement', *BioRxiv*. Cold Spring Harbor Laboratory, p. 214262.

Ge, L. *et al.* (2016) 'Robust 3D Hand Pose Estimation in Single Depth Images: From Single-View CNN to Multi-View CNNs', *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016-Decem, pp. 3593–3601. doi: 10.1109/CVPR.2016.391.

Gentner, R. and Classen, J. (2009) 'Development and evaluation of a low-cost sensor glove for assessment of human finger movements in neurophysiological settings', *Journal of Neuroscience Methods*, 178(1), pp. 138–147. doi: 10.1016/j.jneumeth.2008.11.005.

Goodale, M. A. *et al.* (1991) 'A neurological dissociation between perceiving objects and grasping them', *Nature*, pp. 154–156. doi: 10.1038/349154a0.

Goodale, M. A. *et al.* (1994) 'Separate neural pathways for the visual analysis of object shape in perception and prehension', *Current Biology*, 4(7), pp. 604–610. doi: 10.1016/S0960-9822(00)00132-9.

Goodale, M. A. (2014) 'How (and why) the visual control of action differs from visual perception', *Proceedings of the Royal Society B: Biological Sciences*, 281(1785). doi: 10.1098/rspb.2014.0337.

Goodale, M. A. and Milner, A. D. (1992) 'Separate visual pathways for perception and action', *Trends in Neurosciences*, 15, pp. 20–25.

Grastyán, E. *et al.* (1959) 'Hippocampal electrical activity during the development of conditioned reflexes', *Electroencephalography and Clinical Neurophysiology*, 11(3), pp. 409–430. doi: 10.1016/0013-4694(59)90040-9.

Gray, C. M. *et al.* (1989) 'Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties.', *Nature*, pp. 334–337. doi: 10.1038/338334a0.

Gray, C. M. and Singer, W. (1989) 'Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex.', *Proceedings of the National Academy of Sciences of the United States of America*, 86(5), pp. 1698–1702. doi: 10.1073/pnas.86.5.1698.

Gregoriou, G. G. *et al.* (2009) 'High-Frequency, long-range coupling between prefrontal and visual cortex during attention', *Science*, 324(5931), pp. 1207–1210. doi: 10.1126/science.1171402.

Harris, K. D. *et al.* (2003) 'Organization of cell assemblies in the hippocampus', *Nature*, 424(6948), pp. 552–556. doi: 10.1038/nature01834.

Harris, K. D. *et al.* (2016) 'Improving data quality in neuronal population recordings', *Nature Neuroscience*, 19(9), pp. 1165–1174. doi: 10.1038/nn.4365.

He, K. *et al.* (2016) 'Deep residual learning for image recognition', in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778.

Hearne, L. J. *et al.* (2017) 'Reconfiguration of brain network architectures between resting state and complexity-dependent cognitive reasoning', *The Journal of Neuroscience*, 37(35), pp. 0485–17. doi: 10.1523/JNEUROSCI.0485-17.2017.

Hebb, D. O. (1949) 'The organization of behavior'. John Wiley and Sons.

van den Heuvel, M. P. and Sporns, O. (2013) 'Network hubs in the human brain',

Trends in Cognitive Sciences. Elsevier Ltd, 17(12), pp. 683–696. doi: 10.1016/j.tics.2013.09.012.

Hochreiter, S. (1991) 'Untersuchungen zu dynamischen neuronalen Netzen', *Unpublished doctoral dissertation, Institut für Informatik, Technische Universität, Munchen*, pp. 1–71. Available at: http://people.idsia.ch/~juergen/SeppHochreiter1991ThesisAdvisorSchmidhuber.pdf.

Holden, E., Owens, R. and Roy, G. G. (1999) *3D Hand Tracker for Visual Sign Recognition*, *Technical Report, Department of Computer Science, University of Western Australia*.

Hong, G. and Lieber, C. M. (2019) 'Novel electrode technologies for neural recordings', *Nature Reviews Neuroscience*. Springer US, 20(6), pp. 330–345. doi: 10.1038/s41583-019-0140-6.

Hubel, D. H. (1957) 'Tungsten Microelectrode for Recording from Single Units', *Science*, 125(3247), pp. 549 LP – 550. doi: 10.1126/science.125.3247.549.

Insafutdinov, E. *et al.* (2016) 'Deepercut: A deeper, stronger, and faster multi-person pose estimation model', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9910 LNCS, pp. 34–50. doi: 10.1007/978-3-319-46466-4_3.

Iwai, Y. *et al.* (1996) 'Gesture recognition by using colored gloves', in *1996 IEEE International Conference on Systems, Man and Cybernetics. Information Intelligence and Systems (Cat. No. 96CH35929)*. IEEE, pp. 76–81.

Janssen, P. and Scherberger, H. (2015) 'Visual Guidance in Control of Grasping', *Annual Review of Neuroscience*. Annual Reviews, 38(1), pp. 69–86. doi: 10.1146/annurev-neuro-071714-034028.

Jasper, H. and Penfield, W. (1949) 'Electrocorticograms in man: Effect of voluntary movement upon the electrical activity of the precentral gyrus', *Archiv für Psychiatrie und Nervenkrankheiten*, 183(1–2), pp. 163–174. doi: 10.1007/BF01062488.

Jones, L. A. and Lederman, S. J. (2006) *Human hand function*. Oxford University Press.

Joo, H. *et al.* (2019) 'Panoptic Studio: A Massively Multiview System for Social Interaction Capture', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(1), pp. 190–204. doi: 10.1109/TPAMI.2017.2782743.

Jouhanneau, J. S. *et al.* (2015) 'In Vivo Monosynaptic Excitatory Transmission

between Layer 2 Cortical Pyramidal Neurons', *Cell Reports*, 13(10), pp. 2098–2106. doi: 10.1016/j.celrep.2015.11.011.

Jun, J. J. *et al.* (2017) 'Fully integrated silicon probes for high-density recording of neural activity', *Nature*. Nature Publishing Group, 551(7679), pp. 232–236. doi: 10.1038/nature24636.

Kandel, E. R. *et al.* (2000) *Principles of neural science*. McGraw-hill New York.

Kaufman, M. T. *et al.* (2014) 'Cortical activity in the null space: permitting preparation without movement.', *Nature neuroscience*. Nature Publishing Group, 17(3), pp. 440–8. doi: 10.1038/nn.3643.

Khambhati, A. N., Mattar, M. G. and Bassett, D. S. (2017) 'Beyond modularity : Fine-scale mechanisms and rules for brain network reconfiguration'.

Kohn, A. *et al.* (2020) 'Principles of Corticocortical Communication: Proposed Schemes and Design Considerations', *Trends in Neurosciences*. Elsevier Ltd, 43(9), pp. 725–737. doi: 10.1016/j.tins.2020.07.001.

König, P., Engel, A. K. and Singer, W. (1995) 'Relation between oscillatory activity and long-range synchronization in cat visual cortex', *Proceedings of the National Academy of Sciences of the United States of America*, 92(1), pp. 290–294. doi: 10.1073/pnas.92.1.290.

Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2012) 'Imagenet classification with deep convolutional neural networks', in *Advances in neural information processing systems*, pp. 1097–1105.

Lachaux, J. P. *et al.* (1999) 'Measuring phase synchrony in brain signals', *Human Brain Mapping*, 8(4), pp. 194–208. doi: 10.1002/(SICI)1097-0193(1999)8:4<194::AID-HBM4>3.0.CO;2-C.

LeCun, Y. A. *et al.* (2012) 'Efficient backprop', in *Neural networks: Tricks of the trade*. Springer, pp. 9–48.

LeCun, Y., Bengio, Y. and Hinton, G. (2015) 'Deep learning', *Nature*, 521(7553), pp. 436–444. doi: 10.1038/nature14539.

Lehmann, S. J. and Scherberger, H. (2013) 'Reach and gaze representations in macaque parietal and premotor grasp areas', *Journal of Neuroscience*, 33(16), pp. 7038–7049. doi: 10.1523/JNEUROSCI.5568-12.2013.

Lemon, R. N. (2008) 'Descending pathways in motor control', *Annual Review of Neuroscience*, 31(Cm), pp. 195–218. doi: 10.1146/annurev.neuro.31.060407.125547.

Leonardi, N. *et al.* (2013) 'Principal components of functional connectivity: A new approach to study dynamic brain connectivity during rest', *NeuroImage*. Elsevier Inc., 83, pp. 937–950. doi: 10.1016/j.neuroimage.2013.07.019.

Lepage, K. Q. *et al.* (2013) 'A procedure for testing across-condition rhythmic spike-field association change', *Journal of Neuroscience Methods*. Elsevier B.V., 213(1), pp. 43–62. doi: 10.1016/j.jneumeth.2012.10.010.

Lepage, K. Q., Kramer, M. a and Eden, U. T. (2011) 'The dependence of spike field coherence on expected intensity.', *Neural computation*, 23(9), pp. 2209–2241. doi: 10.1162/NECO_a_00169.

Logothetis, N. K. *et al.* (2001) 'A neurophysiological investigation of the basis of the BOLD signal in fMRI', *Nature*, 412(6843), pp. 150–157. Available at: http://www.nature.com/nature/journal/v412/n6843/full/412150a0.html.

Logothetis, N. K., Kayser, C. and Oeltermann, A. (2007) 'In Vivo Measurement of Cortical Impedance Spectrum in Monkeys: Implications for Signal Propagation', *Neuron*, 55(5), pp. 809–823. doi: 10.1016/j.neuron.2007.07.027.

Luppino, G. *et al.* (1999) 'Largely segregated parietofrontal connections linking rostral intraparietal cortex (areas AIP and VIP) and the ventral premotor cortex (areas F5 and F4)', *Experimental Brain Research*, 128(1–2), pp. 181–187. doi: 10.1007/s002210050833.

Luppino, G. and Rizzolatti, G. (2000) 'The organization of the frontal motor cortex', *News in Physiological Sciences*, 15(5), pp. 219–224. doi: 10.1152/physiologyonline.2000.15.5.219.

MacKenzie, I. S. (1995) 'Input devices and interaction techniques for advanced computing', *Virtual environments and advanced interface design*. Oxford Univ. Press, pp. 437–470.

von der Malsburg, C. (1981) 'The correlation theory of brain function (Internal Report 81-2)', *Goettingen: Department of Neurobiology, Max Planck Intitute for Biophysical Chemistry*.

von der Malsburg, C. and Schneider, W. (1986) 'A neural cocktail-party processor', *Biological Cybernetics*, 54(1), pp. 29–40. doi: 10.1007/BF00337113.

Martínez-Vázquez, P. and Gail, A. (2018) 'Directed Interaction Between Monkey Premotor and Posterior Parietal Cortex During Motor-Goal Retrieval from Working Memory', *Cerebral Cortex*, 28(5), pp. 1866–1881. doi: 10.1093/cercor/bhy035.

Mathis, A. *et al.* (2018) 'DeepLabCut: markerless pose estimation of user-defined body

parts with deep learning', *Nature Neuroscience*. Springer US, 21(September). doi: 10.1038/s41593-018-0209-y.

Michaels, J. A. *et al.* (2020) 'A modular neural network model of grasp movement generation', *bioRxiv*, p. 742189. doi: 10.1101/742189.

Michaels, J. A., Dann, B. and Scherberger, H. (2016) 'Neural Population Dynamics during Reaching Are Better Explained by a Dynamical System than Representational Tuning', *PLoS Computational Biology*, 12(11), pp. 1–22. doi: 10.1371/journal.pcbi.1005175.

Michaels, J. A. and Scherberger, H. (2018) 'Population coding of grasp and laterality-related information in the macaque fronto-parietal network', *Scientific Reports*, 8(1), pp. 1–15. doi: 10.1038/s41598-018-20051-7.

Millett, D. (2001) 'Hans Berger: From psychic energy to the EEG', *Perspectives in biology and medicine*. The Johns Hopkins University Press, 44(4), pp. 522–542.

Mitra, P. P. and Pesaran, B. (1999) 'Analysis of dynamic brain imaging data.', *Biophysical journal*. Elsevier, 76(2), pp. 691–708. doi: 10.1016/S0006-3495(99)77236-X.

Murata, A. *et al.* (1997) 'Object representation in the ventral premotor cortex (Area F5) of the monkey', *Journal of Neurophysiology*, 78(4), pp. 2226–2230. doi: 10.1152/jn.1997.78.4.2226.

Murata, A. *et al.* (2000) 'Selectivity for the shape, size, and orientation of objects for grasping in neurons of monkey parietal area AIP', *Journal of Neurophysiology*, 83(5), pp. 2580–2601. doi: 10.1152/jn.2000.83.5.2580.

Murthy, V. N. and Fetz, E. E. (1996) 'Synchronization of neurons during local field potential oscillations in sensorimotor cortex of awake monkeys.', *Journal of neurophysiology*, 76(6), pp. 3968–3982. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8985893.

Musallam, S. *et al.* (2007) 'A floating metal microelectrode array for chronic implantation', *Journal of Neuroscience Methods*, 160(1), pp. 122–127. doi: 10.1016/j.jneumeth.2006.09.005.

Nácher, V. *et al.* (2013) 'Coherent delta-band oscillations between cortical areas correlate with decision making', *Proceedings of the National Academy of Sciences*, 110(37), pp. 15085–15090. doi: 10.1073/pnas.1314681110.

Nath, T. *et al.* (2019) 'Using DeepLabCut for 3D markerless pose estimation across species and behaviors', *Nature Protocols*, 14(7), pp. 2152–2176. doi: 10.1038/s41596-019-0176-0.

Neher, E. and Sakmann, B. (1992) 'The patch clamp technique', *Scientific American*. JSTOR, 266(3), pp. 44–51.

Nigam, S. *et al.* (2016) 'Rich-Club Organization in Effective Connectivity among Cortical Neurons', *Journal of Neuroscience*, 36(3), pp. 670–684. doi: 10.1523/JNEUROSCI.2177-15.2016.

Nolte, G. *et al.* (2004) 'Identifying true brain interaction from EEG data using the imaginary part of coherency', *Clinical Neurophysiology*, 115(10), pp. 2292–2307. doi: 10.1016/j.clinph.2004.04.029.

Norton, S. and Jewett, R. E. (1965) 'Frequencies of slow potential oscillations in the cortex of cats', *Electroencephalography and Clinical Neurophysiology*, 19(4), pp. 377–386. doi: 10.1016/0013-4694(65)90162-8.

Novak, P., Lepicovska, V. and Dostalek, C. (1992) 'Periodic amplitude modulation of EEG', *Neuroscience Letters*, 136(2), pp. 213–215. doi: 10.1016/0304-3940(92)90051-8.

O'Shea, D. J. *et al.* (2017) 'The need for calcium imaging in nonhuman primates: New motor neuroscience and brain-machine interfaces', *Experimental Neurology*. The Authors, 287, pp. 437–451. doi: 10.1016/j.expneurol.2016.08.003.

Obien, M. E. J. *et al.* (2015) 'Revealing neuronal function through microelectrode array recordings', *Frontiers in Neuroscience*, 9(JAN), p. 423. doi: 10.3389/fnins.2014.00423.

Onorato, I. *et al.* (2020) 'A distinct class of bursting neurons with strong gamma synchronization and stimulus selectivity in monkey V1', *Neuron*. Elsevier Inc., 105(1), pp. 180–197. doi: 10.1016/j.neuron.2019.09.039.

Overduin, S. A. *et al.* (2010) 'An instrumented glove for small primates', *Journal of Neuroscience Methods*. Elsevier B.V., 187(1), pp. 100–104. doi: 10.1016/j.jneumeth.2009.12.007.

Palmigiano, A. *et al.* (2017) 'Flexible information routing by transient synchrony', *Nature Neuroscience*, 20(7), pp. 1014–1022. doi: 10.1038/nn.4569.

Pandarinath, C. *et al.* (2018) 'Inferring single-trial neural population dynamics using sequential auto-encoders', *Nature Methods*. Springer US, 15(10), pp. 805–815. doi: 10.1038/s41592-018-0109-9.

Penttonen, M. and Buzsáki, G. (2003) 'Natural logarithmic relationship between brain oscillators', *Thalamus and Related Systems*, 2(2), pp. 145–152. doi: 10.1016/S1472-9288(03)00007-4.

Pesaran, B. *et al.* (2002) 'Temporal structure in neuronal activity during working memory in macaque parietal cortex', *Nature Neuroscience*, 5(8), pp. 805–811. doi: 10.1038/nn890.

Pesaran, B. (2010) 'Neural correlations, decisions and actions', *Current opinion in neurobiology*, 20(2), pp. 166–171. doi: 10.1016/j.conb.2010.03.003.

Pesaran, B. *et al.* (2018) 'Investigating large-scale brain dynamics using field potential recordings: Analysis and interpretation', *Nature Neuroscience*. Springer US, 21(7), pp. 903–919. doi: 10.1038/s41593-018-0171-8.

Pesaran, B., Nelson, M. J. and Andersen, R. A. (2008) 'Free choice activates a decision circuit between frontal and parietal cortex.', *Nature*, 453(7193), pp. 406–409. doi: 10.1038/nature06849.

Poulet, J. F. A. and Petersen, C. C. H. (2008) 'Internal brain state regulates membrane potential synchrony in barrel cortex of behaving mice', *Nature*, 454(7206), pp. 881–885. doi: 10.1038/nature07150.

Quiroga, R. Q., Nadasdy, Z. and Ben-Shaul, Y. (2004) 'Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering', *Neural computation*. MIT Press, 16(8), pp. 1661–1687.

Rathelot, J. A. and Strick, P. L. (2009) 'Subdivisions of primary motor cortex based on cortico-motoneuronal cells', *Proceedings of the National Academy of Sciences of the United States of America*, 106(3), pp. 918–923. doi: 10.1073/pnas.0808362106.

Regan, D. (1968) 'A high frequency mechanism which underlies visual evoked potentials', *Electroencephalography and Clinical Neurophysiology*, 25(3), pp. 231–237. doi: 10.1016/0013-4694(68)90020-5.

Rehg, J. M. and Kanade, T. (1994a) 'Digiteyes: vision-based hand tracking for human-computer interaction', *Motion of Non-Rigid and Articulated Obgects Workshop, Proceedings*, pp. 16–22. doi: 10.1109/mnrao.1994.346260.

Rehg, J. M. and Kanade, T. (1994b) 'Visual tracking of high DOF articulated structures: An application to human hand tracking', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 801 LNCS(May), pp. 35–46. doi: 10.1007/bfb0028333.

Rizzolatti, G. *et al.* (1988) 'Functional organization of inferior area 6 in the macaque monkey - II. Area F5 and the control of distal movements', *Experimental Brain Research*, 71(3), pp. 491–507. doi: 10.1007/BF00248742.

Rizzolatti, G. and Luppino, G. (2001) 'The cortical motor system', *Neuron*. Elsevier,

31(6), pp. 889–901.

Roelfsema, P. R., Lamme, V. A. F. and Spekreijse, H. (2004) 'Synchrony and covariation of firing rates in the primary visual cortex during contour grouping', *Nature Neuroscience*, 7(9), pp. 982–991. doi: 10.1038/nn1304.

Rossant, C. *et al.* (2016) 'Spike sorting for large, dense electrode arrays', *Nature Neuroscience*, 19(4), pp. 634–641. doi: 10.1038/nn.4268.

Rousche, P. J. and Normann, R. A. (1998) 'Chronic recording capability of the utah intracortical electrode array in cat sensory cortex', *Journal of Neuroscience Methods*, 82(1), pp. 1–15. doi: 10.1016/S0165-0270(98)00031-4.

Rouse, A. G. and Schieber, M. H. (2015) 'Spatiotemporal distribution of location and object effects in reach-to-grasp kinematics', *Journal of Neurophysiology*, 114(6), pp. 3268–3282. doi: 10.1152/jn.00686.2015.

Rumelhart, D. E., Hinton, G. E. and Williams, R. J. (1986) 'Learning representations by back-propagating errors', *Nature*, 323(6088), pp. 533–536. doi: 10.1038/323533a0.

Russakovsky, O. *et al.* (2015) 'ImageNet Large Scale Visual Recognition Challenge', *International Journal of Computer Vision*, 115(3), pp. 211–252. doi: 10.1007/s11263-015-0816-y.

Schaffelhofer, S. and Scherberger, H. (2012) 'A new method of accurate hand- and arm-tracking for small primates', *Journal of Neural Engineering*, 9(2). doi: 10.1088/1741-2560/9/2/026025.

Schaffelhofer, S. and Scherberger, H. (2016) 'Object vision to hand action in macaque parietal , premotor , and motor cortices', *eLife*, 5, pp. 1–24. doi: 10.7554/eLife.15278.

Scherberger, H., Jarvis, M. R. and Andersen, R. A. (2005) 'Cortical local field potential encodes movement intentions in the posterior parietal cortex', *Neuron*, 46(2), pp. 347–354. doi: 10.1016/j.neuron.2005.03.004.

Schieber, M. H. and Hibbard, L. S. (1993) 'How somatotopic is the motor cortex hand area?', *Science*, 261(5120), pp. 489–492. doi: 10.1126/science.8332915.

Schmidhuber, J. (2015) 'Deep Learning in neural networks: An overview', *Neural Networks*, 61, pp. 85–117. doi: 10.1016/j.neunet.2014.09.003.

Schneider, M. *et al.* (2020) 'A general theory of coherence between brain areas', *bioRxiv*, p. 2020.06.17.156190. doi: 10.1101/2020.06.17.156190.

Schwarz, R. J. and Taylor, C. L. (1955) 'The anatomy and mechanics of the human

hand', *Artificial limbs*, 2(2), pp. 22–35.

Sejnowski, T. J., Churchland, P. S. and Movshon, J. A. (2014) 'Putting big data to good use in neuroscience', *Nature Neuroscience*. Nature Publishing Group, 17(11), pp. 1440–1441. doi: 10.1038/nn.3839.

Seth, A. K. (2010) 'A MATLAB toolbox for Granger causal connectivity analysis', *Journal of Neuroscience Methods*, 186(2), pp. 262–273. doi: 10.1016/j.jneumeth.2009.11.020.

Shadlen, M. N. and Newsome, W. T. (1998) 'The Variable Discharge of Cortical Neurons: Implications for Connectivity, Computation, and Information Coding', *The Journal of Neuroscience*, 18(10), pp. 3870–3896. doi: https://doi.org/10.1523/JNEUROSCI.18-10-03870.1998.

Sharp, T. *et al.* (2015) 'Accurate, robust, and flexible realtime hand tracking', *Conference on Human Factors in Computing Systems - Proceedings*, 2015-April, pp. 3633–3642. doi: 10.1145/2702123.2702179.

Siapas, A. G., Lubenov, E. V. and Wilson, M. A. (2005) 'Prefrontal phase locking to hippocampal theta oscillations', *Neuron*, 46(1), pp. 141–151. doi: 10.1016/j.neuron.2005.02.028.

Siegel, M., Warden, M. R. and Miller, E. K. (2009) 'Phase-dependent neuronal coding of objects in short-term memory.', *Proceedings of the National Academy of Sciences of the United States of America*, 106(50), pp. 21341–21346. doi: 10.1073/pnas.0908193106.

Simon, T. *et al.* (2017) 'Hand keypoint detection in single images using multiview bootstrapping', *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017-Janua, pp. 4645–4653. doi: 10.1109/CVPR.2017.494.

Simone, L. K. *et al.* (2007) 'A low cost instrumented glove for extended monitoring and functional hand assessment', *Journal of Neuroscience Methods*, 160(2), pp. 335–348. doi: 10.1016/j.jneumeth.2006.09.021.

Simone, L. K. and Kamper, D. G. (2005) 'Design considerations for a wearable monitor to measure finger posture', *Journal of NeuroEngineering and Rehabilitation*, 2, pp. 1–10. doi: 10.1186/1743-0003-2-5.

Singer, W. (1993) 'Neuronal representations, assemblies and temporal coherence', *Progress in Brain Research*, 95(C), pp. 461–474. doi: 10.1016/S0079-6123(08)60388-X.

Singer, W. and Gray, C. M. (1995) 'Visual Feature Integration and the Temporal

Correlation Hypothesis', *Annual Review of Neuroscience*, 18(1), pp. 555–586. doi: 10.1146/annurev.ne.18.030195.003011.

Sirota, A. *et al.* (2008) 'Entrainment of Neocortical Neurons and Gamma Oscillations by the Hippocampal Theta Rhythm', *Neuron*, 60(4), pp. 683–697. doi: 10.1016/j.neuron.2008.09.014.

Sridhar, S. *et al.* (2015) 'Fast and robust hand tracking using detection-guided optimization', *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 07-12-June, pp. 3213–3221. doi: 10.1109/CVPR.2015.7298941.

Sturman, D. J. and Zeltzer, D. (1994) 'A Survey of Glove-based Input', *IEEE Computer Graphics and Applications*, 14(1), pp. 30–39.

Sussillo, D. *et al.* (2015) 'A neural network that finds a naturalistic solution for the production of muscle activity.', *Nature Neuroscience*, 18(7), pp. 1025–1033. doi: 10.1038/nn.4042.

Taira, M. *et al.* (1990) 'Parietal cortex neurons of the monkey related to the visual guidance of hand movement', *Experimental Brain Research*, 83(1), pp. 29–36. doi: 10.1007/BF00232190.

Thompson, W. H., Brantefors, P. and Fransson, P. (2017) 'From static to temporal network theory: Applications to functional brain connectivity', *Network Neuroscience*, 1(2), pp. 69–99. doi: 10.1162/NETN_a_00011.

Tompson, J. *et al.* (2014) 'Real-time continuous pose recovery of human hands using convolutional networks', *ACM Transactions on Graphics*, 33(5). doi: 10.1145/2629500.

Vargas-Irwin, C. E. *et al.* (2010) 'Decoding complete reach and grasp actions from local primary motor cortex populations.', *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 30(29), pp. 9659–9669. doi: 10.1523/JNEUROSCI.5443-09.2010.

Vinck, M. *et al.* (2010) 'The pairwise phase consistency: A bias-free measure of rhythmic neuronal synchronization', *NeuroImage*. Elsevier Inc., 51(1), pp. 112–122. doi: 10.1016/j.neuroimage.2010.01.073.

Vinck, M. *et al.* (2011) 'An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias', *NeuroImage*. Elsevier Inc., 55(4), pp. 1548–1565. doi: 10.1016/j.neuroimage.2011.01.055.

Vinck, M. *et al.* (2012) 'Improved measures of phase-coupling between spikes and the

Local Field Potential', *Journal of Computational Neuroscience*, 33(1), pp. 53–75. doi: 10.1007/s10827-011-0374-4.

Vinck, M. *et al.* (2013) 'Attentional Modulation of Cell-Class-Specific Gamma-Band Synchronization in Awake Monkey Area V4', *Neuron*. Elsevier Inc., 80(4), pp. 1077–1089. doi: 10.1016/j.neuron.2013.08.019.

De Vittorio, M., Martiradonna, L. and Assad, J. (2014) 'Nanotechnology and neuroscience: Nano-electronic, photonic and mechanical neuronal interfacing', *Nanotechnology and Neuroscience: Nano-Electronic, Photonic and Mechanical Neuronal Interfacing*, pp. 1–285. doi: 10.1007/978-1-4899-8038-0.

Vyas, S. *et al.* (2020) 'Computation Through Neural Population Dynamics', *Annual Review of Neuroscience*, (In press.), pp. 249–275.

Waldert, S., Lemon, R. N. and Kraskov, A. (2013) 'Influence of spiking activity on cortical local field potentials.', *J Physiol*, 591(Pt 21), pp. 5291–303. doi: 10.1113/jphysiol.2013.258228.

Webster, M. J., Bachevalier, J. and Ungerleider, L. G. (1995) 'Transient subcortical connections of inferior temporal areas TE and TEO in infant macaque monkeys', *Journal of Comparative Neurology*, 352(2), pp. 213–226. doi: 10.1002/cne.903520205.

Wei, S.-E. *et al.* (2016) 'Convolutional pose machines', in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pp. 4724–4732.

Wise, K. D., Angell, J. B. and Starr, A. (1970) 'An Integrated-Circuit Approach to Extracellular Microelectrodes', *IEEE Transactions on Biomedical Engineering*, BME-17(3), pp. 238–247. doi: 10.1109/TBME.1970.4502738.

Womelsdorf, T. *et al.* (2006) 'Gamma-band synchronization in visual cortex predicts speed of change detection.', *Nature*, 439(7077), pp. 733–736. doi: 10.1038/nature04258.

Womelsdorf, T., Westendorff, S. and Ardid, S. (2013) 'Subnetwork selection in deep cortical layers is mediated by beta-oscillation dependent firing.', *Frontiers in systems neuroscience*, 7(June), p. 25. doi: 10.3389/fnsys.2013.00025.

Wong, Y. T. *et al.* (2016) 'Coherent neuronal ensembles are rapidly recruited when making a look-reach decision', *Nature Neuroscience*. 2016/01/11. Nature Publishing Group, 19(January), pp. 1–11. doi: 10.1038/nn.4210.

Yun, M. H. *et al.* (1997) 'An instrumented glove for grasp specification in virtual-reality-based point-and-direct telerobotics', *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 27(5), pp. 835–846. doi:

10.1109/3477.623236.

Zhang, F. *et al.* (2020) 'MediaPipe Hands: On-device Real-time Hand Tracking'. Available at: http://arxiv.org/abs/2006.10214.

Zimmerman, T. G. (1985) 'US Patent No. 4542291'.

Zimmerman, T. and Lanier, J. (1991) 'US Patent No 4,988,981'.

5 Appendix A : A general theory of coherence between brain areas

Marius Schneider[1,2], Benjamin Dann[3], Swathi Sheshadri[3,4], Hansjörg Scherberger[3,4], Martin Vinck[1]

[1]Ernst Strüngmann Institute (ESI) for Neuroscience in Cooperation with Max Planck Society, Frankfurt am Main, Germany
[2]International Max Planck Research School for Neural Circuits, Frankfurt am Main, Germany
[3]German Primate Center, Kellnerweg 4, 37077 Göttingen, Germany
[4]Faculty of Biology and Psychology, University of Göttingen, 37073 Göttingen, Germany

[†]Corresponding author. Email: martin.vinck@esi-frankfurt.de

Manuscript is submitted for publication

Preprint DOI: https://doi.org/10.1101/2020.06.17.156190

Contribution: In addition to the projects described in the chapters of this thesis, I assisted with this project titled 'A general theory of coherence between brain areas'. The preprint of this manuscript is included in this appendix. I collaborated with Marius Schneider, Martin Vinck, Benjamin Dann and Hansjörg Scherberger by developing data pre-processing scripts for this project.

A general theory of coherence between brain areas

Marius Schneider[a,d], Benjamin Dann[b], Swathi Sheshadri[b,c], Hansjörg Scherberger[b,c], Martin Vinck[a,e]

[a] *Ernst Strüngmann Institute (ESI) for Neuroscience in Cooperation with Max Planck Society, Frankfurt am Main, Germany*
[b] *German Primate Center, Göttingen, Germany*
[c] *Faculty of Biology and Psychology, University of Goettingen, Germany*
[d] *International Max Planck Research School for Neural Circuits, Frankfurt am Main, Germany*
[e] *Correspondence to: martin.vinck@esi-frankfurt.de*

Abstract

What does neuronal coherence tell us about neuronal communication? Does coherence between field potentials (e.g. LFP, EEG, MEG) reflect spiking entrainment or coupling between oscillators? Is it a mechanism for communication between brain areas, or a byproduct of interareal connectivity? We hypothesized that interareal coherence is explained by the fact that outputs from one cortical area give rise to synaptic inputs in the same brain area, and correlated synaptic inputs in another area. Our mathematical analysis demonstrates that coherence between a sending and receiving area is precisely predicted from only two parameters: Interareal connectivity and oscillatory synchronization in the sending area. This model predicts narrow-band coherence even in case of a flat transfer function and in the absence of spiking entrainment in a downstream area, and reproduces frequency-specific Granger-causality patterns between brain areas (gamma feedforward, beta feedback). In general, we find that Granger-causality between field potentials is dominated by oscillatory synchronization in the sending area, whereas spiking entrainment is determined by the resonant properties of the receiver. Our model accurately reproduces LFP-LFP beta-coherence between macaque areas 7B and F5 in the absence of beta phase-locking within area F5. Together, these findings uncover a precise mechanistic model of interareal coherence as a (by)product of connectivity and power.

Keywords: Coherence; oscillation; synchronization; gamma; beta; connectomics; connectivity; entrainment; Wilson-Cowan; 7B; F5; macaque; resonance; feedforward; feedback; spike-field coherence; PPC (pairwise phase consistency); Granger-causality; LFP (Local Field Potential); EEG (Electroencephalography); ECoG (Electrocorticogram); MEG (Magnetoencephalography).

Introduction

The brain is a dynamical system that generates intelligent behavior through the interaction between different brain areas (Buzsáki, 2006; Miller and Wilson, 2008; Varela et al., 2001; Bressler, 1995; Engel et al., 2001; Singer and Gray, 1995; Nicolelis et al., 1995; Siegle et al., 2019; Fries, 2015). These interareal interactions can be studied by measuring temporal correlations (e.g. coherence, Granger-causality, cross-correlations) between electrophysiological signals from multiple brain areas. Electrophysiological signals are commonly distinguished into spike recordings and measures of population synaptic activity, for example LFP, ECoG, EEG, MEG (Nunez and Srinivasan, 2006; Pesaran et al., 2018; Buzsáki et al., 2012; Mitzdorf, 1985; Einevoll et al., 2013); we refer to the latter as "field potentials". Field potentials have obvious disadvantages as compared to spike recordings, such as: (i) The loss of spatial resolution, (ii) volume conduction, and (iii) the fact that synaptic potentials are a mixture of local and afferent inputs (Pesaran et al., 2018; Nunez and Srinivasan, 2006; Buzsáki and Schomburg, 2015; Einevoll et al., 2013; Buzsáki et al., 2012). Yet, they also have clear advantages: (i) They can be recorded non-invasively or from the cortical surface; (ii) by pooling over synaptic potentials in a large cortical volume, they can uncover weak interactions between areas.

Field potentials from different brain areas show coherent activity in various frequency bands (Buzsáki, 2006). Interareal coherence is influenced by several cognitive and behavioral factors (Grothe et al., 2012a; Gregoriou et al.,

2009; Salazar et al., 2012a; Richter et al., 2018; Colgin et al., 2009; Buschman and Miller, 2007; Fries, 2015; Varela et al., 2001; Babapoor-Farrokhran et al., 2017; Phillips et al., 2014; Olcese et al., 2016; Montgomery and Buzsáki, 2007; Von Stein and Sarnthein, 2000; Bressler et al., 1993; Brunet et al., 2014). Furthermore, distinct frequency bands are thought to play specific roles in interareal communication. For example, gamma (30-80Hz) and alpha/beta frequencies (10-30Hz) have been related to feedforward and feedback corticocortical communication, respectively (Buschman and Miller, 2007; Bastos et al., 2015; van Kerkoerle et al., 2014; Richter et al., 2018; Bressler et al., 2006; Mejias et al., 2016). Yet, the unequivocal functional significance and causal interpretation of these findings remains to be established. Does interareal coherence itself have an influence on the communication between areas? Or is interareal coherence a byproduct of connectivity, and hence communication, between areas (coherence through communication)?

The interpretation of interareal coherence between field potentials is fraught with many problems (Pesaran et al., 2018; Buzsáki and Schomburg, 2015; Nolte et al., 2004; Nunez and Srinivasan, 2006; Einevoll et al., 2013; Vinck et al., 2015, 2010). A well-known problem is the spread of electromagnetic fields over space (volume conduction) (Sirota et al., 2008a; Nunez and Srinivasan, 2006; Pesaran et al., 2018; Vinck et al., 2016; Carmichael et al., 2019; Parabucki and Lampl, 2017). In this paper, we investigate another major problem which we refer to as the *synaptic mixing problem*: In the normal LFP range (<80Hz), field potentials primarily reflect the summed synaptic activity (transmembrane currents) in a population of neurons (Einevoll et al., 2013; Pesaran et al., 2018; Nunez and Srinivasan, 2006; Buzsáki et al., 2012). These synaptic potentials can be decomposed into two parts: (i) Synaptic inputs caused by spikes from neurons in the *same* brain area, and (ii) afferent synaptic inputs caused by spikes from neurons in *other* brain areas. Likewise, spiking activity in one brain area (A) can cause synaptic potentials in the *same* brain area (A), and highly correlated synaptic potentials in *another* brain area (B) at a delay. We refer to these effects as "synaptic mixing". As a consequence, electric signals measured in area A and area B may, in part, be delayed copies of the same underlying signal, which would trivially give rise to interareal coherence and Granger-causality (Pesaran et al., 2018; Buzsáki and Schomburg, 2015). Because the transmission of the signal in this case is not instantaneous, the synaptic mixing problem cannot be solved with techniques that address the volume conduction problem (Trongnetrpunya et al., 2016; Pesaran et al., 2018; Nolte et al., 2004; Haufe et al., 2012; Vinck et al., 2015, 2011).

Here, we develop a general theory of the way in which synaptic mixing determines interareal coherence, using mathematical analysis, simulations of neuronal populations, and analysis of interareal recordings.

Results

Beta-coherence between areas F5 and 7B

We start out with the analysis of neural data, in which two distant brain areas show clear beta-synchronization between LFP signals. Beta-synchronization is thought to be involved in motor preparation, maintenance of a cognitive state and top-down modulation (Richter et al., 2018; Buschman and Miller, 2007; Salazar et al., 2012a; Bastos et al., 2015; Engel et al., 2001; Scherberger et al., 2005). We recorded from subdivisions of the parietal (area 7B) and premotor (area F5) cortex. These brain areas are involved in tasks like the reaching and grasping of objects (Dann et al., 2016). Premotor area F5 is one of the main projection targets of area 7B; this is a strong long-range projection, as area F5 lies several cm's away from area 7B (Johnson et al., 1996; Luppino et al., 1999; Markov et al., 2014).

We recorded LFPs and spiking activity using two 32-channel flexible microelectrode arrays per area (Figure 1A). We analyzed the memory period of the task (see Methods), in which beta oscillations are most prominent. We observed a clear beta-peak (≈20 Hz) in the power spectrum of 7B LFPs (Figure 1B). To analyze how single units were synchronized with LFPs, we computed the unbiased spike-field PPC value, which is proportional to the squared spike-field coherence (Vinck et al., 2012). Consistent with the clear LFP beta-peak, single units in 7B showed significant spike-LFP phase-locking in the beta-frequency band (Figure 1C). Spike-LFP locking was close to zero for frequencies outside the beta-band (Figure 1C). We also found coherence between the two 7B electrode grids (Figure 1F,H). Thus, synaptic and spiking activity in area 7B showed clear beta-band synchronization.

Next, we examined interareal LFP-LFP and spike-LFP coherence between area 7B and area F5. We observed relatively strong and narrow-band beta-coherence between 7B and F5 LFPs (Figure 1E). This would suggest, *prima facie*, oscillatory coupling between these two brain areas. To analyze whether this coherence was due to volume conduction, we computed a measure of coherence that avoids spurious coherence due to volume conduction, namely the Weighted

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.17.156190. this version posted June 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

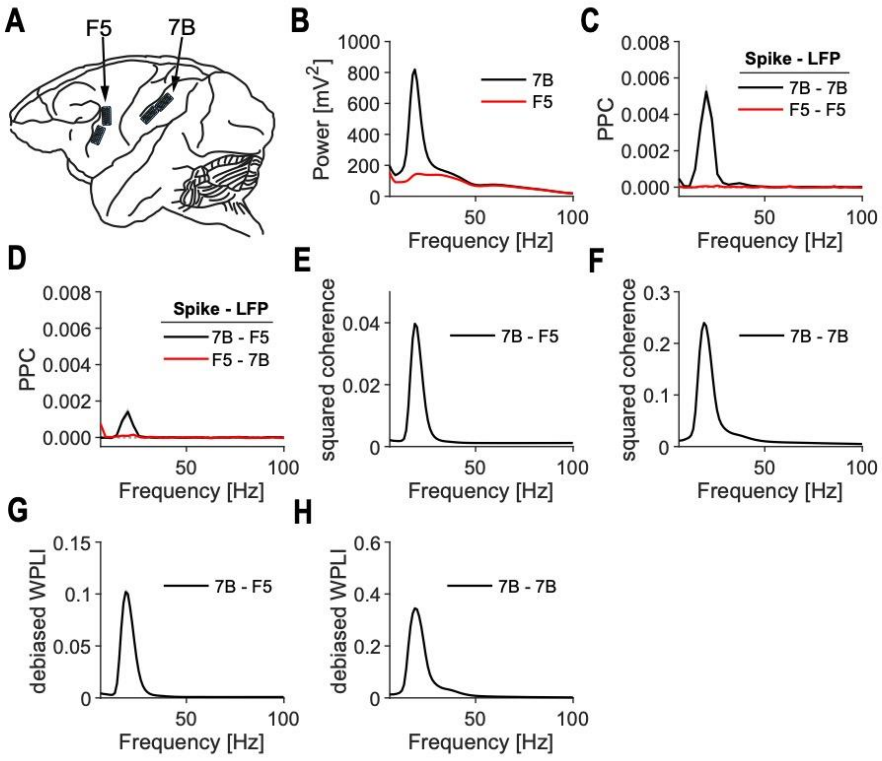

Figure 1: Coherence without spiking entrainment between macaque F5 and 7B **(A)** Illustration of electrode FMA grid recordings from the premotor areas F5 and 7B of a macaque monkey during grasping task (see Methods). We analyzed the memory period before motor execution. **(B)** LFP power spectrum for 7B (black) and F5 (red). **(C)** Spike-field phase locking (measured with PPC; see Methods) for different combinations of spikes and LFPs: Spikes in 7B to LFPs in 7B (black); spikes in F5 to LFPs in F5 (red). Phase locking values were averaged over all electrodes in corresponding grid. Spikes were pooled across all neurons in a session. The average number of neurons per session were 22.3 (7B) and 17.9 (F5). **(D)** Same as (C), but now between spikes in 7B to LFPs in F5 (black), and spikes in F5 to LFPs in 7B (red). **(E-F)** Coherence between medial and lateral 7B LFPs (F) and 7B and F5 LFPs (E). **(G-H)** Absolute value of weighted phase-lag index (Vinck et al., 2011) between 7B and F5 LFPs, as well as medial 7B and lateral 7B LFPs. The debiased WPLI is a measure of phase-synchronization robust to volume conductionn (Vinck et al., 2011). All figures have standard errors of the mean. For PPC the standard error is across 4 conditions × 6 recording sessions. For coherence and WPLI, 6 sessions were pooled for the computation, and the standard error was computed across 4 conditions.

Phase Lag Index (WPLI; see Methods) (Vinck et al., 2011). The WPLI spectrum showed beta-synchronization between 7B and F5 LFPs, suggesting that LFP-LFP coherence was not due to volume conduction (Figure 1G); this is consistent with the large spatial distance between 7B and F5.

Because there was clear beta-coherence between 7B and F5 LFPs, we expected to also find interareal beta-synchronization between spikes and LFPs. Surprisingly, we did not detect significant beta-band spike-LFP phase-locking between F5 spikes and F5 or 7B LFPs (Figure 1C-D). Moreover, we found that the LFP power spectrum in area F5 was dominated by the $1/f$ component, and showed only a small peak in the beta-band (Figure 1B). Thus, we found clear evidence for beta-oscillatory activity in area 7B and beta-coherence between F5 and 7B LFPs, but no beta-synchronization within F5. How can this discrepancy be explained?

Coherence predicted from connectivity and power

In this section, we will show that field-field (e.g. LFP-LFP, EEG-EEG or MEG-MEG) coherence between two brain areas ("Areas 1 and 2") can be predicted from two parameters: (i) The connection strength between Areas 1 and 2; (ii) the structure of the LFP power spectrum in both areas. We start out with the simple case of unidirectional communication, where Area-1 projects to Area-2 with a connection weight w (Figure 2A). In this case, output spikes from Area-1 can cause synaptic potentials both in Area-1 (through recurrent connections) and in Area-2 (through interareal projections). At the same time, spikes from neurons in Area-2 will also cause synaptic potentials in Area-2. Thus, the LFP in Area-2 will be a mixture of synaptic inputs from Area-1 and Area-2. We model the measured signal in Area-1 as the sum of an oscillatory process and a broad-band process, e.g. $1/f$ pink noise. We further suppose that that the intrinsic signal of Area-2 has no rhythmic (oscillatory) component. The Area-2 LFP is therefore described as a linear mixture of its own background fluctuations and the synaptic inputs from Area-1.

To understand the behavior of coherence in this model, we generated synthetic signals in both areas. A beta-oscillation in Area-1 was generated using a dampened harmonic oscillator (AR(2); see Methods). The background processes were generated as $1/f$ pink-noise signals (see Methods). As expected, Area-1 showed a clear beta-frequency peak in the power spectrum (Figure 2A). However, because of the small connection weight ($w = 0.069$), there were no visible oscillations in the power spectrum of Area-2. Moreover, the transfer function from Area-1 to Area-2 was flat, reflecting the linear superposition of inputs from Area-1 onto Area-2, and the lack of any form of filtering. Despite this flat transfer function, LFP-LFP coherence and Granger-causality between Area-1 and 2 showed a clear spectral peak in the beta frequency-band. Thus, a narrow-band peak in the coherence spectrum emerged as a byproduct of synaptic mixing.

To generalize the results of this simulation, we derived an analytical expression for the coherence spectrum based on the connection weight w and the SOS ("Sender Oscillation Strength"). The SOS was denoted α, defined as the ratio of the power spectral density of the oscillatory component, $S_{11}(f)$, relative to the background signal, $H_{11}(f)$. The squared-magnitude coherence (approximately equal to Granger-causality) between Area-1 and Area-2 equals (see Methods)

$$C^2(f) = \Theta\left(w^2\left(1+\alpha(f)\right)\right) \approx w^2\left(1+\alpha(f)\right). \tag{1}$$

Here $\Theta(x) \equiv \frac{x}{1+x}$ is a sigmoid-like function (Figure 2B). We confirm this analytical expression with numerical simulations based on generating AR(2) and $1/f$ signals (Figure 2C). This analytical expression has two main implications:

(1) Coherence and Granger-Geweke causality are monotonically increasing functions of the SOS $\alpha(f)$ and the coupling weight w. The coherence peaks at the frequency where $\alpha(f)$ reaches a maximum. Even for small values of w, coherence can show a prominent peak if $\alpha(f)$ is large. In this model, band-limited coherence between field potentials is a byproduct of communication: The coherence "itself" does not contribute to communication, because the transfer function is flat and there is no interaction between the inputs from Area-1 and the intrinsic activity of Area-2.

(2) The dependence of coherence on coupling weight has a non-linear, sigmoidal form. Hence, a change in coupling weight w causes a steep change in coherence for some values of w, but weakly affects the coherence for other values of w. Specifically, the derivative reaches a maximum for $w = 1/\sqrt{1+\alpha}$. Thus, when $\alpha(f)$ is large, steep changes in coherence already occur for relatively small coupling weights. In general, the effect of the coupling weight w on the coherence depends on the value of the SOS, $\alpha(f)$. Furthermore, a change in the coherence can be caused both by change in interareal connectivity and the SOS (Figure 2D).

Modelling the coherence between 7B and F5

We used the synaptic mixing model of coherence to reproduce the LFP-LFP coherence between areas 7B and F5. To this end, we generated beta oscillations in Area-1 (7B) as dampened harmonic oscillators, and the background processes as $1/f^n$ spectra (see Methods). This model produced an LFP-LFP coherence spectrum with a clear peak in the beta-frequency band. However, it overestimated the LFP-LFP coherence at other frequencies (Figure 2E). This suggests that our simple model of coherence cannot fully explain the interareal coherence between 7B and F5. Could there still be some "genuine" oscillatory coupling between areas 7B and F5?

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.17.156190. this version posted June 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

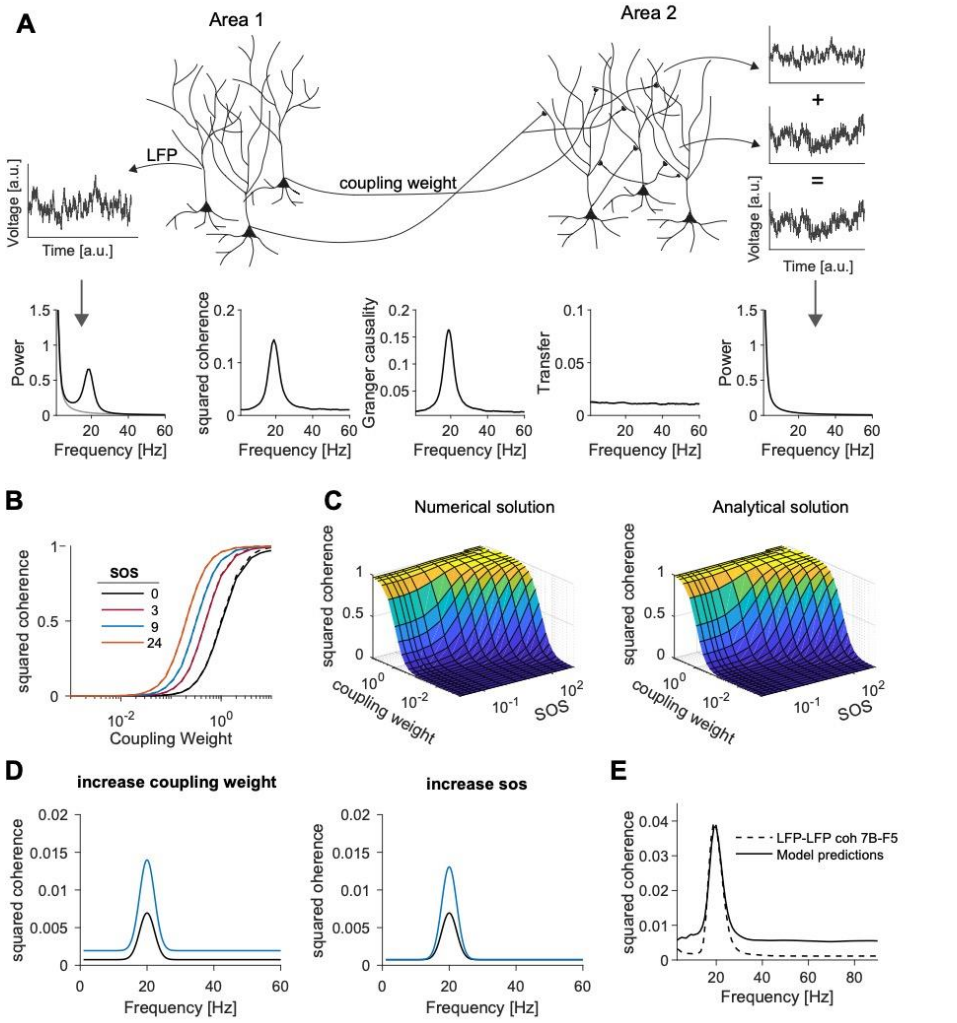

Figure 2: Interareal coherence as a (by)product of connectivity and power. **(A)** Illustration of the *synaptic mixing problem*. The Area-2 LFP is a linear superposition of intrinsic activity and afferent inputs from Area 1 weighted by w. The signal in Area-1 was modelled as a superposition of an oscillatory process and a $1/f$ background signal. The intrinsic activity in Area-2 was modelled by the same background signal. The power spectrum in Area-1 but not Area-2 shows a clear beta-peak. The coherence and Granger-causality spectra show clear beta peaks, but the transfer function is flat. The SOS (Sender Oscillation Strength) at the oscillatory frequency $f_1 = 20Hz$ was $SOS(f_1) = 14$; $w = 0.1$. **(B-C)** Coherence as a function of the SOS and coupling weight. Analytical derivation matched the numerical simulations (C), performed using autoregressive models with varying oscillation strengths. **(D)** Coherence spectra for two "behavioral" conditions, in which either the coupling weight (left) or the SOS (right) changed. The change in coherence was greater at the sender's oscillation frequency. The parameters were: $SOS = 10$ (blue and black - left), $w = 0.025$ (black - left), $w = 0.035$ (blue - left) and $SOS = 10$ (black - left), $SOS = 20$ (blue - left), $w = 0.025$. **(E)** Modelling interareal coherence between areas 7B and F5. Power spectra were fitted as a linear mixture of an AR(2) model with $1/f^n$ background fluctuations ($w = 0.069$). The coherence at the peak frequency can be well reproduced, but that the coherence at other frequencies is not fitted well.

The influence of the projection-source coherence

In this Section, we uncover an additional mechanism that further suppresses the coherence related to non-oscillatory background fluctuations, namely the *projection-source coherence* (Figure 3A); the influence of the projection-source coherence makes the synaptic mixing problem fundamentally different from the volume conduction problem. In the model analyzed above, the background fluctuations in the sender were projected onto the receiver *in the same way* as the oscillatory process in the sender. Specifically, the signal projected from Area-1 onto Area-2 was modelled as a fully coherent copy of the signal in Area-1. In other words, we assumed that the *inputs* into Area-2, caused by Area-1 activity, were a fully coherent copy of the total activity in Area 1. However, because of the anatomy of interareal connections, this assumption will be wrong:

The Area-1-to-2 projections will originate from a relatively small subset of neurons in Area-1, because some neurons only project to other brain areas, and some neurons project only locally (Markov et al., 2011, 2014; Han et al., 2018; Lur et al., 2016). Because the inputs into Area-2 are only caused by a *subset* of neurons in Area-1, the inputs into Area-2 will be only be partially correlated with the *total* activity in Area-1. In other words, the coherence between the summed activity of the Area-1-to-Area-2 projection neurons and the Area-1 LFP may not be equal to 1. We refer to this as the *projection-source coherence*. We will show that this coherence is an increasing function of two factors: (i) The fraction of projecting neurons. (ii) The spike-LFP coherence of the Area-1-to-Area-2 projecting neurons with the Area-1 LFP. In general, we expect that oscillations substantially enhance the spike-field coherence of individual neurons (Onorato et al., 2020; Sirota et al., 2008b; Vinck et al., 2016; Buffalo et al., 2011; Chalk et al., 2010; Buzsáki and Schomburg, 2015; Gregoriou et al., 2009) A related point, is that the projecting area typically consists of spatially separated populations of projecting neurons. The outputs of these different populations may converge onto a single receiver. In the absence of coherent activity, the LFP tends to be highly local, spanning a small cortical volume ($\approx$ 200μm) (Lindén et al., 2011; Einevoll et al., 2013; Katzner et al., 2009). In such a volume, the total number of neurons will be only around 160000/125 = 1280 neurons (for a density of 160000 neurons/mm^3) (Christensen et al., 2007). Out of these 1280 neurons, only few to tens of active neurons might be projecting to another given brain area. However, the coherence across populations can be strongly enhanced by rhythmic synchronization (Figure 1) (Buzsáki, 2006; Gray et al., 1989). This effectively increases the number of coherently firing projection neurons, and the spatial reach of the LFP (Lindén et al., 2011).

In the Methods section, we derive a general expression for the coherence between Area-1 and Area-2, which contains an additional dependence on the projection-source coherence, denoted $C_{\text{proj,source}}(f)$:

$$C^2(f) \approx w^2\,(\alpha(f)+1)\,C^2_{\text{proj,source}}(f)\,. \tag{2}$$

Here $\alpha(f)$ is the ratio of intrinsic power in the sender over intrinsic power in the receiver. This equation states that the original influence of the SOS and w on coherence is multiplied by the projection-source coherence. Thus, oscillations in the sender increase the interareal coherence because of two factors, $C_{\text{proj,source}}(f)$ and $\alpha(f)$.

Next, we derive an expression for the projection-source coherence $C_{\text{proj,source}}(f)$ (see Methods) as

$$C^2_{\text{proj,source}} \approx N_p\,\phi^2(f)\ . \tag{3}$$

Here, $\phi^2(f)$ is the squared spike-LFP coherence of an individual projecting neuron with the Area-1 LFP. The variable N_p represents the number of Area-1-to-Area-2 projecting neurons. By combining all equations we state our main analytical result, namely that

$$C^2 \approx w^2\,(\alpha+1)\ N_p\,\phi^2(f). \tag{4}$$

Note that the same equation applies to the coherence between Area-1-LFP and Area-2 spiking, if Area-2 spiking relates in a linear or sigmoidal way to synaptic inputs (see Methods). In sum, LFP-LFP coherence is determined by the following four factors:

1. The connection weight w *relative* to other inputs into the receiver.
2. The number of active projecting neurons in Area-1.
3. The spike-field coherence of projection neurons. From neural recordings, the value $\phi^2(f)$ can be estimated using the unbiased spike-field PPC, which approximates the spike-field coherence (Vinck et al., 2012; Onorato et al., 2020; Vinck et al., 2016).

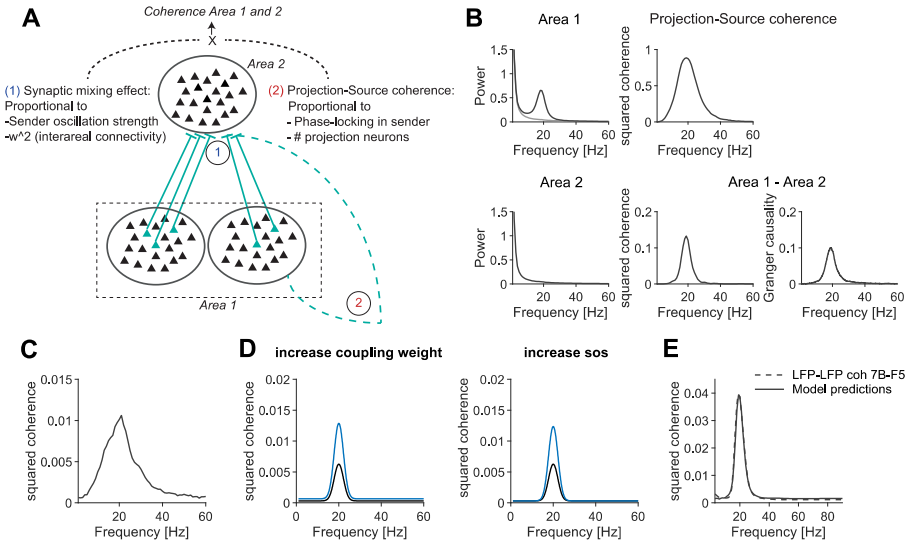

Figure 3: Interareal coherence is strongest at the frequency where the sender exhibits oscillatory synchronization. **(A)** Illustration of two mechanisms due to which coherence is suppressed at frequencies where the sender does not exhibit synchronization. The sending area consists out of multiple local sub-populations, whose projections may converge onto another area. We expect the high coherence at the oscillatory frequency, but low coherence for the background $1/f$ fluctuations. In addition, the sending area contains only a small population of Area-1-to-2 projecting neurons. The summed potential of these projecting neurons will be mostly coherent with the Area-1 LFP at the oscillation frequency. **(B)** Simulation where the subset of Area-1-to-2 projecting neurons is most coherent with the Area-1 LFP at the oscillation frequency (Top, right). The interareal coherence and Granger-causality (Bottom) show spectral peaks only at the oscillation frequency. In this case, the SOS at the oscillatory frequency was $SOS = 14$ and coupling weight $w = 0.1$; the oscillation was modelled as an AR(2). The background fluctuations in Area-1 were only partially transmitted, with a weight of $\sqrt{1-\gamma}$, $\gamma = 1$ (see Methods). **(C)** Same as in Figure (B), but now with the presence of an oscillation in Area-2 with the same strength as Area-1. In this case the coherence is substantially lower than in (B), because the SOS equals 1 for all frequencies. However, it still exhibits a spectral peak because of the coherence of the Area-1 LFP with the Area-1-to-2 projection neurons. **(D)** Both increases in coupling weight and SOS cause a narrow-band increase in interareal coherence. The parameters were: $SOS = 10$ (blue and black - left), $w = 0.025$ (black - left), $w = 0.036$ (blue - left) and $SOS = 10$ (black - right), $SOS = 20$ (blue - right), $w = 0.025$. **(E)** LFP-LFP coherence between macaque 7B and F5 (dashed) can be reproduced (solid) by the synaptic mixing model shown in (B). The background fluctuations in Area-1 are only partially transmitted, with a weight of $\sqrt{1-\gamma}$, $\gamma = 0.95$, $w = 0.077$.

4. The sender-oscillation-strength $\alpha(f)$.

We further observe that the SOS and the spike-LFP PPC are proportional to each other, i.e. $\alpha(f) \propto \phi^2(f)$ (see Methods). For small values of N_p, $\alpha(f)$ also depends linearly on N_p (see Methods). Moreover, the number of projecting neurons in an area should be proportional to the connection weight w between the areas (Markov et al., 2011, 2014). Hence we obtain the supra-linear relationship

$$C^2(f) \propto w^4\, \phi^4(f) \tag{5}$$

for small values of w and ϕ. Hence, 2-fold changes in spike-LFP coherence or firing rates in the sending area can, *ceteris paribus*, cause a 16-fold change in the squared interareal coherence. For larger values of w and ϕ, the expression takes the linear form $C^2(f) \propto w^2\, \phi^2(f)$ as $C_{\text{proj,source}}$ is bound by 1.

The connection weight w depends on several factors: (i) The number of synaptic connections that are made into another area; (ii) factors modulating synaptic efficacy, like compartementalized dendritic inhibition and neuromodulators (Batista-Brito et al., 2018; McGinley et al., 2015; Chiu et al., 2013).; (iii) the termination zone of the synapses, given that synaptic currents on basal and apical dendrites cancel each other out (Lindén et al., 2011); (iv) the average firing rates of the projecting neurons.

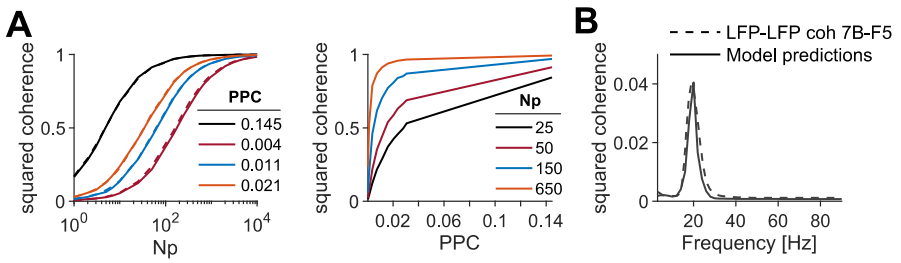

Figure 4: Analytical derivation of projection-source coherence, and interareal coherence predicted through spike-LFP Phase Locking in sender area **(A)** Squared coherence between Area-1 LFP and the summed activity of Area-1-to-2 projection neurons, as a function of the number of projecting neurons (Np) and the phase-locking of individual neurons (spike-field PPC). For this simulation, we generated an AR(2) signal in the sender and generated spikes in 10000 neurons according to inhomogenous Poisson processes modulated by this AR(2) process (see Methods). The dashed lines show a tight match between our analytical derivations and the simulations (see Methods). **(B)** Interareal coherence spectrum between 7B and F5. Data: dashes. Model: solid. Model predictions were inferred from: the spike-LFP PPC within area 7B; the SOS $\alpha(f)$; and the coupling weight w of the model in Figure 2 (E). The total number of Neurons in area 7B is 100000 (which was arbitrarily chosen) and 550 of these are projecting to area F5.

Given the dependence of interareal coherence on so many factors, it is highly non-trivial to infer changes in interareal synaptic gain as a function of cognition or behavior. To make matters worse, controlling for the average firing rate or spike-field coherence will be insufficient, because projection neurons are a highly specialized subclass of cells (Lur et al., 2016; Han et al., 2018).

Simulations of extended model

We performed several simulations to validate our theoretical model. The projection-source coherence can be incorporated by simulating the activity in the sender as a weighted superposition of two $1/f$ background processes (see Methods) (Figure 3B): The projected $1/f$ background had a weight of $\sqrt{0.05}$, and the non-projected background a weight of $\sqrt{0.95}$. In this case, the projection-source coherence equals 0.25 for the $1/f$ background process. We found that, as expected, the coherence was strongly suppressed at frequencies outside the oscillation frequency band.

An important consequence of the analytical expression in Eq. 2 is that a peak in the coherence spectrum can emerge even when the sender and receiving area have identical oscillation strength, and are not phase-synchronized. By contrast, the simple synaptic-mixing model would not predict a peak in the coherence spectrum, because $\alpha(f)$ is now equal to 1 for all frequencies. However, due to the dependence of the interareal coherence on the projection-source coherence, the coherence does attain a narrow-band structure (Figure 3C). Changes in coherence with the SOS $\alpha(f)$ or interareal connectivity w also occurred in a narrow frequency range, and were only visible at the oscillation frequency of the sender (Figure 3D).

Finally, we performed additional simulations to confirm the analytical expression obtained for the projection-source coherence (Eq. 3, 33, 39) (Figure 4A).

Explaining coherence between areas 7B and F5

We found that the coherence spectrum between 7B and F5 LFPs could not be accurately reproduced based on the SOS and interareal connectivity alone. We therefore applied the extended model, in which the $1/f$ background is only partially transmitted. The extended model was able to accurately reproduce the observed LFP-LFP coherence between 7B and F5 (Figure 3E). We further modelled the interareal coherence based on the spike-LFP PPC within area 7B (Figure 4B). For this, we used the analytical expression for the coherence based on spike-LFP PPC, the SOS $\alpha(f)$, coupling weight w and the number of projecting neurons N_p. The coherence was well predicted from this model: The required number of projecting neurons to reproduce the coherence was about 550. Note that the literature reports about two orders of magnitude more 7B-to-F5 projection neurons for a given retrograde injection in F5 (Markov et al., 2011, 2014). Together, these data suggest that beta coherence between F5 and 7B LFPs can be explained by the synaptic mixing model.

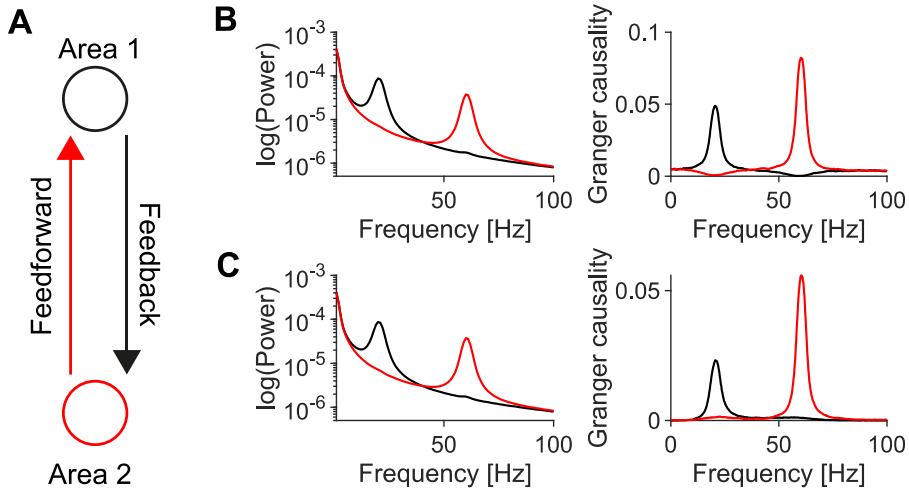

Figure 5: Gamma bottom-up and beta top-down Granger-causality are byproducts of connectivity and differences in power spectra. **(A)** We model feedforward (bottom-up) and feedback (top-down) interactions between two areas. The intrinsic activity of each area is modelled as a linear mixture of an AR(2) model and $1/f^n$ background fluctuations. The AR(2) model of Area-1 is oscillating at 20 Hz. The AR(2) model of Area-2 is oscillating at 60 Hz. The signals are transmitted to the other area with a conduction delay of 5ms. **(B)** We simulate the interareal interactions according to the simple linear synaptic mixing model shown in Figure 2. Shown from left to right are: Power spectra of Area-1 and 2 and Granger-causality in feedforward (red) and feedback (black) direction. Thus, feedforward Granger at gamma and feedback Granger at beta frequencies can be reproduced by synaptic mixing alone, but some distortions are observed for feedback gamma and feedforward beta. **(C)** Same as (B), but now for the model where the background fluctuations in Area-1 are only partially transmitted, with a weight of $\sqrt{1-\gamma}$, $\gamma = 0.95$. This again gives rise to gamma feedforward and beta feedback, but flattens the Granger-causality spectrum at frequencies outside the oscillation bands, producing a better fit to the published literature (Bastos et al., 2015). In (B) and (C) the signals were transmitted with a weight of $w = 0.08$.

Explaining frequency-dependent interactions (gamma bottom-up, beta top-down)

In the models above, we considered a unidirectional communication setting. In reality, brain areas are typically bidirectionally connected (Markov et al., 2011). The extent to which activity in Area-1 predicts activity in Area-2, and vice versa, can be quantified using Granger-Geweke-causality (Geweke, 1982). Previous studies in primate visual and parietal areas have suggested that *feedforward* and *feedback* Granger-causality are respectively strong at gamma and alpha/beta frequencies (Bastos et al., 2015; van Kerkoerle et al., 2014; Buschman and Miller, 2007; Mejias et al., 2016; Michalareas et al., 2016). One possible interpretation of these findings is that brain areas communicate with different frequencies in the feedforward or feedback direction (Bastos et al., 2015). Alternatively, frequency-specific Granger-causality might be a consequence of the presence of distinct oscillation bands in different brain areas, not of frequency-specific transfer functions. In particular, there exists a gradient of oscillation frequencies and time constants across the cortical hierarchy; gamma and beta oscillations are prominent in early visual areas and parietal cortex, respectively (see also Figure 1B) (Figure 5) (Bastos et al., 2015; Brovelli et al., 2004; Bosman et al., 2012; Vinck et al., 2016; Salazar et al., 2012a; Murray et al., 2014; Onorato et al., 2020; Spyropoulos et al., 2020; Scherberger et al., 2005).

We modelled the intrinsic signals in both Area-1 and Area-2 as the sum of a broad-band process (pink noise), and an oscillatory signal (Area-1: gamma; Area-2: beta) (Figure 5A). We then coupled the areas bidirectionally. Our synaptic mixing models accurately reproduced the Granger spectra previously reported, with stronger feedforward and feedback Granger-causality at gamma and beta frequencies, respectively (Figure 5B-C).

E/I networks

We found that narrow-band LFP-LFP coherence results from synaptic mixing alone. However, it is possible that oscillatory synaptic inputs cause some spiking entrainment in the receiver, which could further increase LFP-LFP coherence. Note that with "spiking entrainment", we simply mean that some spikes in the receiver are caused by the sender, not that the entire intrinsic oscillation in a receiver is phase-locked to the oscillation in a sender. To investigate this question, we modelled each area as a network of E/I neurons whose dynamics are governed by stochastic Wilson-Cowan equations. These networks show noisy oscillations that mimick oscillatory behavior in the brain (Spyropoulos et al., 2020; Wallace et al., 2011; Powanwe and Longtin, 2019; Mejias et al., 2016). Note that the E/I network did not contain dendritic low-pass filtering, which would in practice diminish the influence of excitatory to excitatory connections at higher frequencies (Buzsáki and Schomburg, 2015; Pike et al., 2000). We constructed two types of scenarios: *Mixing without entrainment:* In Simulation 1, spikes from Area-1 generated field potentials in Area-2, but the neurons in Area-2 were "blind" to the inputs from Area-2. Thus, there was synaptic mixing, but no spiking entrainment, and LFP-LFP coherence was strictly due to synaptic mixing. *Mixing with entrainment:* In Simulation 2, there was spiking entrainment, i.e. spiking in the receiver was modulated by synaptic inputs from the sender. We simulated different cases, e.g. gamma oscillations in the sender, beta oscillations in the receiver, or a combination of these. In Figure 6, we only show coherence, because Granger-causality is approximately equal to the squared coherence for unidirectional communication (see Methods).

Without spiking entrainment, we found strong LFP-LFP coherence at the frequency of the sender (Figure 6B-C), as predicted from the results obtained above. However, spiking entrainment in the receiver did not substantially contribute to the interareal LFP-LFP coherence if the sender and the receiver had different oscillation frequencies (Figure 6B-C,E).

In Figure 6D, we considered a scenario where the sender and the receiver had the same oscillation frequency (gamma) and power. Without spiking entrainment, we found only weak gamma LFP-LFP coherence 6D. This is due to the fact that the SOS was now matched to the receiver oscillation strength, i.e., $\alpha(f) = 1$ for all f (see also Figure 3). With spiking entrainment, we observed a strong increase in LFP-LFP gamma coherence. For low interareal connectivity, LFP-LFP coherence now mostly reflected spiking entrainment (Figure 6E), in contrast to the case where the receiver oscillated in a different frequency band (Figure 6B,E); this difference reflects the resonant properties of the receiver. Although LFP-LFP coherence was now more strongly influenced by spiking entrainment, the value of LFP-LFP coherence was comparable to the case where the receiver oscillated in the beta band (Figure 6E). Overall, the contribution of spiking entrainment to LFP-LFP coherence increased as the difference in oscillation frequency between sender and receiver became smaller (Figure 6G).

For all cases of sender-receiver frequencies, LFP-LFP coherence showed the stereotypical sigmoidal increase as a function of interareal connection strength. For the gamma-to-gamma simulation, spike-field phase-locking to the sender oscillations was initially weak; this indicated that most spikes in the receiver were triggered by its own intrinsic oscillations, but some spikes were triggered by the afferent oscillatory inputs (Figure 6F). With an increase in interareal connectivity, we observed a gradual increase in spike-field phase-locking (Figure 6F). Thus, there was no sudden phase transition where the intrinsic oscillations in the receiver were fully phase-locked to the sender.

These findings show that communication between a sender and receiver can produce LFP-LFP coherence in two ways: (i) By synaptic mixing, and (ii) by triggering spikes in receiver. In both cases, the amount of coherence showed a similar sigmoidal dependence on the interareal connectivity; the strength of interareal coherence was determined by interareal connectivity strength. Interareal LFP coherence should in both cases also depend on the projection-source coherence and the factors that govern it. Finally, the frequency at which spiking entrainment, i.e. actual communication, will be prominent, is determined by the resonant properties of the receiver, not the sender (Figure 7). However, the frequency at which LFP-LFP coherence or Granger-causality will be large is determined by the oscillatory properties of the sender, not the receiver.

Discussion

We showed that interareal coherence between a sending and receiving area depends on four factors: (i) The spike-field coherence of projection neurons in the sender. (ii) The oscillation strength in the sender relative to the intrinsic activity in the receiver. (iii) The interareal connectivity. (iv) The number of projection neurons. We concluded that

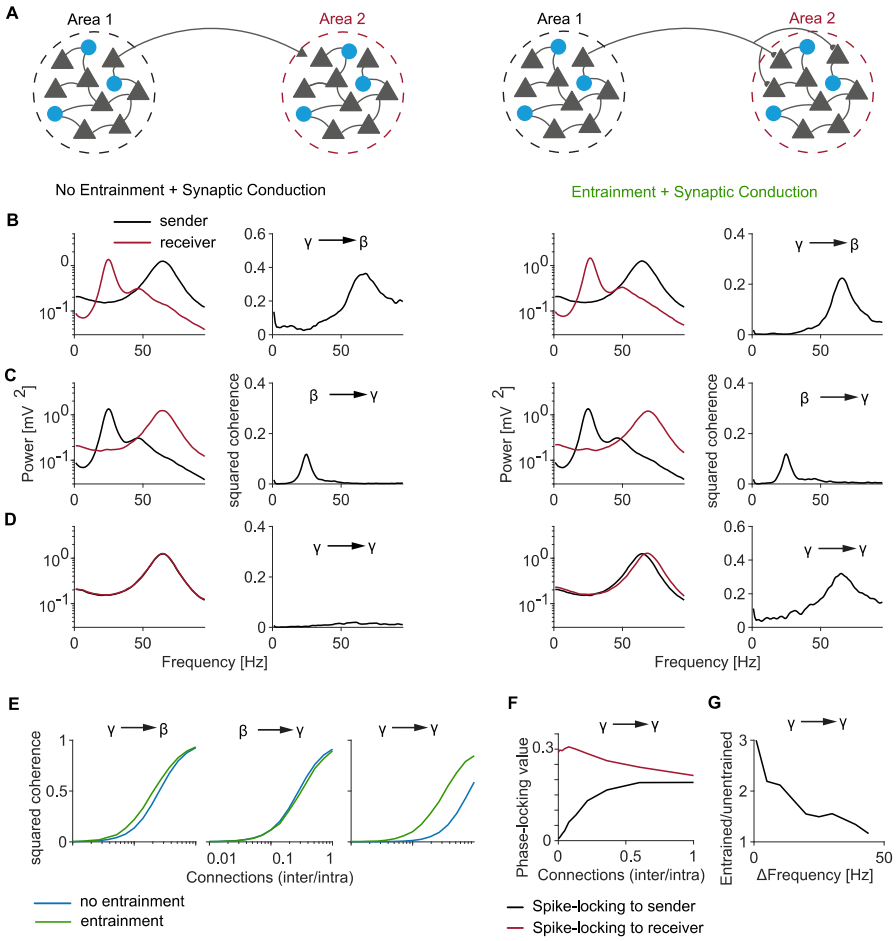

Figure 6: Spiking entrainment in the receiver does not substantially contribute to interareal LFP-LFP coherence, except when the frequencies in the sender and receiver overlap. **(A)** Illustration of the two models. Each area consisted of a population of spiking neurons whose dynamics were modelled by stochastic Wilson-Cowan equations. In the first model (left two columns), synaptic potentials due to inputs from Area-1 were superimposed onto the synaptic potentials from Area-2 itself. Neurons in Area-2 were "blind" to the synaptic inputs from Area-1, i.e. spiking entrainment was prohibited. The second model (right two columns) is identical to the first model, however synaptic inputs from Area-1 could now entrain the neurons in Area-2. **(B)** First two columns: Sender oscillates at gamma and the receiver at beta. Coherence spectra show clear peaks, following the power in the sender. Last two columns: Spiking entrainment increases coherence slightly. **(C)** and **(D)** Same as in (B), but now with different oscillation frequencies. When the oscillation frequency in the sender matches with the receiver, there is a increase in LFP-LFP coherence due to spiking (D). (B-D) all for coupling value of $w = 0.1$. **(E)** Change in coherence as a function of the ratio of inter- to intra-regional connection rates. **(F)** Spike-field phase-locking-value of neurons in the receiver population to the oscillations in the receiver (i.e. sum of all synaptic inputs caused by spikes in Area 2) and oscillations in the sender (i.e. sum of all synaptic inputs caused by spikes in Area 1). As the number of connections increases, the phase locking gradually increases. **(G)** Change in coherence going from unentrained to entrained case, expressed as a ratio, for $w = 0.05$. A value of 3 means 3 times more coherence in the entrained than the unentrained case.

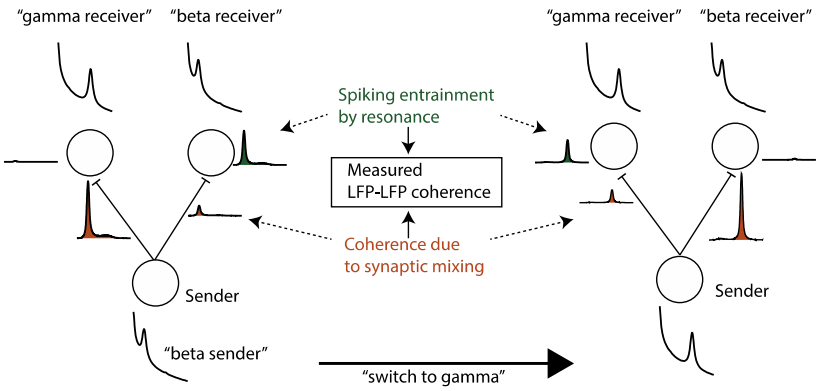

Figure 7: Illustration of difference between LFP-LFP coherence and spiking entrainment, and ability to switch communication by changing oscillations in the sender. In the left case, the sender oscillates at beta, and it would have a high LFP-LFP coherence with a receiver at gamma due to synaptic mixing. However, this LFP coherence does not translate into spiking entrainment. When the receiver also oscillates at beta, the LFP coherence due to synaptic mixing is lower, but due to resonance, the receiver will now exhibit more spiking entrainment. In the right case, the sender switches to gamma, and now switches communication to the gamma receiver. Paradoxically, LFP-LFP coherence might be higher between the sender and the receiver that communicate less.

narrow-band coherence between field potentials is a predicted byproduct of power and connectivity: It does not require coupling between oscillators, frequency-specific information transmission, or spiking entrainment.

Dynamic changes in interareal coherence with cognition can be caused by a combination of these four factors, including the interareal connectivity. Although the anatomical substrate of interareal connectivity changes slowly, interareal connectivity can be rapidly modulated by several mechanisms. These include (i) compartmentalized dendritic inhibition and (ii) neuromodulators (Batista-Brito et al., 2018; McGinley et al., 2015; Chiu et al., 2013). Gain modulation of firing rates in the sending area also leads to an effective increase in interareal connectivity, if the firing rates of other sources of synaptic input do not change. We conclude that band-limited changes in interareal LFP-LFP coherence do not imply frequency-specific changes in information transmission. In fact, a change in interareal connectivity is expected to increase information transfer at *all* frequencies. However, at the level of LFP-LFP coherence, a change in e.g. interareal connectivity only has *measurable* effects at the frequencies where neurons in the sending area fire synchronously (Figure 3).

A function for coherence?

There are numerous theories ascribing functional roles to interareal phase-synchronization and coherence in coordinating neuronal interactions, e.g. (Varela et al., 2001; Bressler, 1995; Engel et al., 2001; Kreiter, 2006; Fries, 2005; Miller and Wilson, 2008; Abeles, 1982; Bonnefond et al., 2017; Salinas and Sejnowski, 2001; Singer and Gray, 1995; Palmigiano et al., 2017; Börgers and Kopell, 2008; Buzsáki and Draguhn, 2004; Vinck et al., 2013a). Various aspects of these theories have been summarized in the CTC ("communication-through-coherence") hypothesis (Fries, 2005, 2009, 2015), which contains three premises: (i) Interareal coherence reflects phase synchronization between the *intrinsic* oscillations in the sender and receiver. (ii) Interareal communication is enhanced when afferent synaptic inputs consistently arrive at an excitable phase of the intrinsic oscillation (Volgushev et al., 1998; Burchell et al., 1998). (iii) Selective communication is implemented through selective coherence (Fries, 2015).

The current paper takes a very different point of view: Two brain areas can only communicate if they are connected, and if they are connected, they will exhibit coherence at a "good" phase-relationship. This is due to the fact that the sending area will be coherent with the inputs that the sender causes in the receiving area. The resulting coherence

is a consequence of communication, not a cause of it. To demonstrate coherence between *intrinsic oscillations*, it is therefore imperative to rule out that interareal coherence is not due to synaptic mixing. Otherwise, the cause (connectivity and communication) can be easily confused with the effect (coherence). And, if coherence is byproduct of communication, then it is unclear what experimental outcomes would possibly falsify CTC; if cognition is expected to increase interareal communication, e.g. due to attention, then an increase in coherence would *a priori* be an expected outcome.

Theories like CTC also propose that communication between two connected areas can be blocked by the absence of coherence (Fries, 2009): Yet, synaptic mixing models also predict that there is no coherence if there is no communication. Furthermore, communication is not consistently blocked by the absence of coherence, because interareal phase-relationships randomly fluctuate between "good" and "bad" phases (Akam and Kullmann, 2012). Alternatively, interareal communication may be prohibited by interareal coherence with a consistent "bad" phase-relationship (Volgushev et al., 1998; Burchell et al., 1998; Fries, 2005; Tiesinga and Sejnowski, 2010; Akam and Kullmann, 2012). This could block communication quite effectively (Volgushev et al., 1998; Burchell et al., 1998; Tiesinga and Sejnowski, 2010; Akam and Kullmann, 2012) and is not predicted by synaptic mixing models; however, phase separation between competing inputs has not been reported in studies that examined the modulation of interareal coherence by attention (Grothe et al., 2012a; Bosman et al., 2012).

How do we disentangle coherence through communication from phase-synchronization between intrinsic oscillations? The strength of interareal coherence may be one indicator as to whether coherence can be explained by synaptic mixing. Because interareal connections are typically weak (Markov et al., 2014, 2011), synaptic mixing is unlikely to yield very high coherence values; these would be more indicative of a pacemaker or strong coupling between oscillators. However, the most compelling correlational evidence for CTC is a moderate change in V1-V4 gamma-coherence with attention (from about 0.06 to 0.09, i.e. squared-magnitude coherence values below 0.01) (Ferro et al., 2020; Bosman et al., 2012; Grothe et al., 2012b). Notably, area V1 contains a very strong source of narrow-band gamma, which is associated with a unique class of excitatory neurons (Gray and McCormick, 1996; Onorato et al., 2020) and shows up to 300-fold power increases (Spyropoulos et al., 2020)). Thus, it should be easy to induce weak V1-V4 gamma coherence and feedforward Granger with synaptic mixing models (see Figure 5). A local increase in the firing rates and phase-locking of V1 or V2 projection neurons with attention would then be sufficient to increase V1-V4 coherence, even in the absence of coupling between local V1 and V4 oscillators (Luck et al., 1997; van Kerkoerle et al., 2014; Buffalo et al., 2011; Chalk et al., 2010).

These considerations highlight a basic problem, namely how to experimentally identify *intrinsic* oscillations using "local" field potential signals, which contain a mixture of local and afferent synaptic inputs (Buzsáki and Schomburg, 2015; Pesaran et al., 2018; Saleem et al., 2017). The strength and prevalence of oscillations show great variation across the cortical sheet; distinct brain regions have particular spectral profiles (Buzsáki, 2006). A few brain regions exhibit very strong oscillations under specific sensory or behavioral conditions. For example: there is a strong source of gamma in V1/V2 (Gray et al., 1989; Peter et al., 2019; Vinck and Bosman, 2016; Onorato et al., 2020; Henrie and Shapley, 2005; Spyropoulos et al., 2020); beta in parieto-frontal cortex (Figure 1,(Scherberger et al., 2005; Dann et al., 2016; Brovelli et al., 2004; Salazar et al., 2012b; Hagan et al., 2012; Donoghue et al., 1998; Murthy and Fetz, 1996)); and theta and gamma in rodent hippocampus (Buzsáki, 2006; Colgin et al., 2009; Bragin et al., 1995). Coherence between these oscillatory sources and areas with weak or no intrinsic oscillations will be dominated by synaptic mixing (Figure 1, (Schomburg et al., 2014)); due to synaptic mixing, oscillations will now appear in areas without intrinsic oscillations. The contribution of afferent inputs to the LFP might depend strongly on the cortical layer. Feedforward projections target the granular layer 4, which has relatively little recurrent connectivity (Lund et al., 2003), and may not exhibit intrinsic oscillatory activity (Livingstone, 1996; Xing et al., 2012). Local field potentials in layer 4 might therefore be dominated by synaptic mixing of afferent inputs, and spiking entrainment in pyramidal neurons might be strongly diminished because of dendritic low-pass filtering (Buzsáki and Schomburg, 2015). To identify that oscillations have at least some local component, it is critical to analyze local spiking activity. However, because interareal projections may induce only weak spiking entrainment in the receiver, spiking entrainment by itself does not demonstrate intrinsic oscillations in the receiver. Suppression of brain areas can be a useful diagnostic tool: For example, (Saleem et al., 2017) showed that in mouse V1, luminance/locomotion-related LFP oscillations in the 60-65Hz range are driven by LGN afferents and not affected by pharmacological suppression of V1 spiking activity.

Even if there are clear intrinsic oscillations with overlapping frequency bands in two connected areas, it is often unclear what mechanism would make them phase-synchronized, and how to measure this. In principle, intrinsic

oscillations can be phase-synchronized by a pacemaker that entrains multiple areas. Clear examples are subcortical areas like the thalamus (sleep spindles) and the medial septum (hippocampal theta) (Steriade et al., 1993; Wang, 2002). Yet it is unclear whether there are pacemakers for neocortical rhythms like gamma and beta. The notion of a pacemaker becomes especially problematic for oscillations that are generated by local circuits and show substantial, stochastic fluctuations in instantaneous amplitude and frequency (e.g. V1 gamma) (Spyropoulos et al., 2020; Burns et al., 2011); in this case a pacemaker would have to continuously reset the ongoing rhythm.

Previous studies have raised other theoretical concerns about the suitability of coherence as a mechanism for selective communication: (1) How can communication be made selective when there is no intrinsic oscillatory activity? For example, V1 gamma can be found for some (e.g. large gratings) but not for other stimuli (e.g. irregular textures), even though both can be seen and attended to (Hermes et al., 2015; Vinck and Bosman, 2016; Henrie and Shapley, 2005; Peter et al., 2019; Ray and Maunsell, 2015). How does attention change interareal coherence for the latter kind of stimuli? (2) Spectral coherence can only be defined over longer time periods, and is usually computed over many trials, in contrast to instantaneous phase. For reliable, selective transmission of one out of multiple sources of afferent inputs, very strong coherence differences and a long integration window are needed (Akam and Kullmann, 2012). This holds true especially when oscillators show stochastic fluctuations in instantaneous frequency (Akam and Kullmann, 2012).

Connectivity mapping through coherence

How to move forward? The synaptic mixing problem can to some extent be addressed by analyzing the (laminar) current-source-densities instead of field potentials, and techniques like ICA (Pesaran et al., 2018; Mitzdorf, 1985; Einevoll et al., 2013; Buzsáki and Schomburg, 2015; Buzsáki et al., 2012; Schomburg et al., 2014). To interpret LFP-LFP coherence, it is also critical to analyze spike-spike and spike-field synchronization (Pesaran et al., 2018; Buzsáki and Schomburg, 2015). In general, it is not *a priori* obvious whether rhythmicity in weak afferent inputs is at all transferred to spiking outputs, and whether weak entrainment is physiologically meaningful or not. Spiking entrainment may be prevented or boosted by several various factors: (i) Non-linearities in single neurons and recurrent networks; (ii) Dendritic low-pass filtering in pyramidal neurons (Buzsáki and Schomburg, 2015; Pike et al., 2000), which counteracts spiking entrainment in the gamma-frequency range (Schomburg et al., 2014; Vinck et al., 2016; Buzsáki and Schomburg, 2015); (iii) Resonance in the receiver (Figure 6).

Even if neuronal responses are linearly or sigmoidally related to afferent synaptic inputs, spike-field coherence can still be explained as a by-product of interareal connectivity and the projection-source coherence (see Methods). For various reasons, the interareal spike-field coherence of a *single* neuron might be very weak compared to the field-field coherence (see Methods, and Figure 1). One reason is that each neuron in the receiver might be targeted by only a few projection neurons, such that it sees a noisy and distorted copy of the rhythm of the sending area (see Methods).

Despite its limitations, interareal LFP-LFP coherence remains a promising tool for studying interareal connectivity and dynamic changes therein, especially in human ECoG. For example, (Bastos et al., 2015; Michalareas et al., 2016) have shown that in the primate visuo-parietal system, interareal Granger-causality between LFP signals can predict layer-specific anatomical projection patterns. Our results suggest that interareal connectivity can be best predicted at the frequencies where one of the areas has a strong intrinsic oscillation (Figure 3 and 6). Improved estimates can be obtained based on the spike-field coherence of projecting neurons (Figure 3. Further progress can be made by factoring in "third area sources" in connectivity and coherence estimates, and the contribution of spiking entrainment. Thus, the presented theoretical model of coherence and Granger-causality opens new avenues for mapping interareal connectivity in the human brain, providing an interesting alternative to DTI.

Acknowledgements and Authorship contributions

We thank Prof. Dr. Wolf Singer, Dr. Georgios Spyropoulos, Patrick Jendritza, and Dr. Craig Richter for very helpful comments. Conceptualization: MS and MV. Mathematical analysis: MS and MV. Simulations and data analysis: MS. Macaque surgeries, recordings and data preprocessing: BD, SS and HS. Writing: MS and MV. Supervision: MV. This projected was supported by ERC Starting Grant to MV (SPATEMP) and a BMF Grant to MV (Bundesministerium fuer Bildung und Forschung, Computational Life Sciences, project BINDA, 031L0167).

Methods

Subjects

Neural activity was recorded simultaneously from many channels in one female rhesus macaque monkey (Animals S, body weight 9, 7 kg). Detailed experimental procedures have been described previously (Dann et al., 2016). All procedures and animal care were in accordance with German and European law and were in agreement with the Guidelines for the Care and Use of Mammals in Neuroscience and Behavioral Research (National Research Council, 2003).

Macaque data

The monkey was trained to perform a delayed grasping task. In this task, the monkey was either instructed to grasp a target with one of the two possible grip types (power and precision), or free to choose between the grip types, as described in detail in previous studies (Dann et al., 2016). During instructed trials, the monkey was visually cued by one of two discs displayed on a monitor to perform the associated grip type. During free-choice trials, both discs were displayed and monkeys could choose freely between grip type. To encourage switching behavior during consecutive free-choice trials, the reward was iteratively reducing every time the monkey repeatedly chose the same grip type. Note that also delayed instructed trials were part of the task. These were not used in this study and are therefore not further explained. The monkey learned to perform the task with high accuracy of 95 +- 0.01 % SD successful trials on average.

Surgical procedures have been described in detail previously (Dann et al., 2016). In short, the monkey was implanted with four chronically implanted 32-channel microelectrode arrays (FMAs; Microprobes for Life Sciences; 32 electrodes; spacing between electrodes: 0.4mm; length: 1.5 to 7.1 mm monotonically increasing to target grey matter along the sulcus), two in part of the ventral premotor cortex (area F5) and two in area 7B, specifically around the anterior intraparietal area (AIP), for a total of 128 channels. Electrode signals from the implanted arrays were amplified and digitally stored using a 128 channel recording system (Cerebus, Blackrock Microsystems; sampling rate 30 kS/s; 0.6-7500Hz band-pass hardware filter).

To detect spikes, electrode signals were first high-pass filtered with a median filter (window length 3ms) and then low-pass filtered with a non-causal Butterworth filter (5000 Hz; 4 th order). Next, common noise-sources were eliminated by applying principal component (PC) artifact cancellation and spike waveforms were detected and semi-automatically sorted using a modified version of the offline spike sorter Waveclus. Finally, redetection of the different average waveforms (templates) was done to detect overlaid waveforms. The exact procedures of spike detection are described previously (Dann et al., 2016). Note that only well isolated single units were used for all analyses. To detect LFPs, electrode signals were first low-pass filtered with a median filter (window length 6.7 ms) and then high-pass filtered with a non-causal Butterworth filter (1 Hz; 4th order). In order to filter outpower line noise and their harmonic and additional band-stopfilter filtering out signals between 49 and 51 Hz and 98 and 102 was applied. Subsequently, signals were down-sampled by averaging 30 consecutive frames from 30000 to 1000 Hz. Broken channels and trials containing movement noise were removed from all further data analyses. For this purpose, the total power, the correlation and the maximum deflection of all channels and trails was compared and all outliers discarded. Finally, to reduce the influence of the on array ground and reference electrode on each array, the trimmed mean over all channels per array (leaving highest two and the lowest two values per time point out) was removed by using linear regression. After spike and LFP detection, single neuron spike events were binned in non- overlapping 1-ms windows to obtain an equal sampling rate of 1000 Hz for both signals. Subsequently, signals were aligned to cue and movement onset for the instructed- and free- choice-task, since activity was locked to both events.

All analyses of macaque data were performed in Matlab (Mathworks) using custom scripts and the FieldTrip toolbox (Oostenveld et al., 2011). Power and coherence spectra were assessed using integration windows of 0.35s length moved over the whole data in steps of 50ms. The epochs were Hann tapered to avoid spectral leakage. Pairwise phase consistency (PPC) between spikes and LFPs was calculated using windows of 350ms around every spike (Vinck et al., 2012), using the spiketriggeredspectrum functions in the FieldTrip SPIKE toolbox. To compute spike-LFP PPC, we first pooled the activity of single units in the area together, which gives the most sensitive estimate of entrainment in an area by increasing the number of pairwise phase comparisons (Vinck et al., 2013b).

Predicting interareal coherence based on connectivity and power

In this Section we derive an analytical expression for coherence based on interareal connectivity and power.

We start out from a unidirectional communication setting, where brain Area-1 projects to brain Area-2. The measured signals are denoted $z_1(t)$ and $z_2(t)$. In the following derivations, and our simulations, we assume that the signals are measured without the addition of extrinsic noise. That is we assume that all signals reflect neural activity, and we assume that there is no volume conduction.

We model the signal $z_1(t)$ in Area-1 as the sum of an oscillatory process $s_1(t)$ and a broad-band process, e.g. Pink noise, $\eta_1(t)$:

$$z_1(t) \equiv s_1(t) + \eta_1(t)\,. \tag{6}$$

The intrinsic signal $z_2(t)$ of Area-2 has no rhythmic component and is modelled as a linear mixture of its own noise term and the projected input from Area-1,

$$z_2(t) = \eta_2(t) + w\,(s_1(t) + \eta_1(t)) \tag{7}$$

where w denotes the projection strength from Area-1 to Area-2. We assume that the background processes $\eta_1(t)$ and $\eta_2(t+\tau)$ are linearly uncorrelated for all τ. For the purpose of mathematical derivation, we suppose that the power spectral densities of the broad-band processes are equal for all f, i.e. $H_{11}(f) = H_{22}(f) \equiv H(f)$, with f frequency. We denote the spectral density of $s_1(t)$ as $S_{11}(f)$. We define the SOS ("Sender Oscillation Strength") as

$$\alpha(f) \equiv \frac{S_{11}(f)}{H_{11}(f)} \equiv \frac{S_{11}(f)}{H(f)}\,. \tag{8}$$

The cross-spectral density between $z_1(t)$ and $z_2(t)$ equals

$$\begin{aligned} Z_{12}(f) &= w\,(S_{11}(f) + H_{11}(f)) \\ &= w\,Z_{11}(f) \end{aligned} \tag{9}$$

and is real-valued. Note that the other cross-terms fell out because we assumed that η_2, η_1 and s_1 are uncorrelated. The squared coherence $C^2(f)$ between Area-1 and Area-2 is defined by

$$C^2(f) \equiv \frac{|Z_{12}(f)|^2}{Z_{11}(f)\,Z_{22}(f)}. \tag{10}$$

This simplifies as follows:

$$\begin{aligned} C^2(f) &= \frac{w^2 Z_{11}(f)^2}{Z_{11}(f)\,Z_{22}(f)} \\ &= \frac{w^2\,Z_{11}(f)}{Z_{22}(f)}\,. \end{aligned} \tag{11}$$

Since $\eta_2(t)$ and $s_1(t)$ are uncorrelated, we have

$$Z_{22}(f) = H(f) + w^2 Z_{11}(f)\,. \tag{12}$$

Eq. 8 now reduces to

$$\begin{aligned} C^2(f) &= \frac{w^2\,Z_{11}(f)}{H(f) + w^2\,Z_{11}(f)} \\ &= \frac{1}{1 + H(f)\,(w^2\,Z_{11}(f))^{-1}}\,. \end{aligned} \tag{13}$$

From Eq. 8 it follows that

$$Z_{11}(f) = H(f)\,(\alpha(f) + 1)\,. \tag{14}$$

Hence $H_{22}(f)\left(w^2 Z_{11}(f)\right)^{-1}$ reduces to the expression

$$\frac{1}{w^2\left(\alpha(f)+1\right)}\,. \tag{15}$$

Thus the coherence can be simplified to

$$\begin{aligned} C^2(f) &= \frac{w^2\left(1+\alpha(f)\right)}{1+w^2\left(1+\alpha(f)\right)} \\ &= \Theta\left(w^2\left(1+\alpha(f)\right)\right), \end{aligned} \tag{16}$$

where $\Theta \equiv \frac{x}{1+x}$ is the sigmoid function.

We can estimate the connectivity weight from the measurement variables by solving for w and $\alpha(f)$,

$$\begin{aligned} w &= \sqrt{\frac{C^2(f)}{-(\alpha(f)+1)(C^2(f)-1)}} \\ &\approx \sqrt{\frac{C^2(f)}{\alpha(f)+1}}, \end{aligned} \tag{17}$$

Here, the approximation is based on the first-order Taylor expansion of the coherence around $C(f) = 0$. We can also take the Taylor expansion around $w = 0$ for Eq. 16 and obtain

$$C^2(f) \approx w^2(1+\alpha(f))\,. \tag{18}$$

Note that the same model derivations (and the derivations below) pertain to Granger-causality, because for unidirectional coupling the following relationship holds between Geweke-Granger causality and coherence (Geweke, 1982):

$$G_{1\to 2}(f) = -\ln(1-C^2(f)) \approx C^2(f)\,, \tag{19}$$

where the approximation was made based on the first order Taylor-expansion around $C(f) = 0$.

Model of interareal coherence taking into account projection patterns

In the model above we assumed that the signal received by the receiver is fully coherent with the signal in the sender. As explained in more detail in the Results Section, this is likely not the case for two reasons: 1) The sender consists of sub-populations that are not fully coherent with each other, especially for frequencies where there is no oscillatory synchronization. 2) The number of projecting neurons in Area-1 may be small, and the coherence between the summed potential of Area-1-to-2 projection neurons and the Area-1 LFP (the projection-source coherence) may not be 1.

Expression of the coherence based on power, interareal connectivity and coherence between Area-1-to-2 projection neurons and the Area-1 LFP.

We first derive an expression of the interareal coherence that includes a linear dependence on the projection-source coherence.

We model the signals as

$$z_1(t) = s_1(t) + \eta_1(t) \tag{20}$$

$$z_2(t) = \eta_2(t) + w\left(s_1^*(t) + \epsilon_1(t)\right)\,. \tag{21}$$

Here, $s_1^*(t)$ is the projected oscillatory signal into Area-2, and $\epsilon_1(t)$ is the projected background signal into Area-2. The coherence between $\eta_1(t)$ and $\epsilon_1(t)$ is denoted $C_{\eta,\epsilon}(f)$. The coherence between $s_1(t)$ and $s_1^*(t)$ is denoted $C_{s,s}(f)$.

We assume that $s_1(t)$ and $s_1^*(t)$ have the same power spectral densities. Likewise we assume that $\eta_1(t)$, $\eta_2(t)$ and $\epsilon_1(t)$ have the same power spectral densities.

We now obtain

$$\begin{aligned} Z_{12}(f) &= w\left(C_{s,s}(f)\, S_{11}(f) + C_{\eta,\epsilon}(f)\,, H(f)\right) \\ &= w\left(C_{s,s}\,\alpha(f) + C_{\eta,\epsilon}(f)\right) H(f)\,; \\ Z_{11}(f) &= S_{11}(f) + H(f) \\ &= (1+\alpha(f))\, H(f)\,; \\ Z_{22}(f) &= w\, S_{11}(f) + (1+w)\, H(f) \\ &= (1 + w\,(1+\alpha(f)))\, H(f) \end{aligned} \tag{22}$$

The squared coherence $C^2(f)$ now simplifies as

$$C^2(f) \equiv \frac{w^2\,\left(C_{ss}(f)\,\alpha(f) + C_{\eta,\epsilon}(f)\right)^2}{(1+\alpha(f))\,(1+w\,(1+\alpha(f)))}\,. \tag{23}$$

Plugging in $\alpha(f) = 0$ for all f we obtain

$$C_{12}^2(f) = \frac{w^2\, C_{\eta,\epsilon}^2(f)}{w+1} \tag{24}$$

$$\approx w^2\, C_{\eta,\epsilon}^2(f) \tag{25}$$

where the first-order Taylor expansion was made around $w = 0$. Thus, the squared coherence between areas scales with the coupling weight and the squared interareal coherence in the sender. For the oscillatory part, assuming the background fluctuations have coherence close to zero, we have

$$C^2(f) \approx w^2 \frac{\alpha(f)^2}{\alpha(f)+1} C_{s,s}^2(f) \tag{26}$$

Following the same derivation we can also obtain an expression for the squared coherence that combines both the noise and the oscillatory term as

$$\begin{aligned} C^2(f) &\equiv \frac{w^2\,\left(C_{\text{proj,source}}(f)\,(\alpha(f)+1)\right)^2}{(1+\alpha(f))\,(1+w\,(1+\alpha(f)))} \\ &\approx C_{\text{proj,source}}^2\,(\alpha+1)\, w^2\,. \end{aligned} \tag{27}$$

Here $C_{\text{proj,source}}$ is the projection-source coherence, and $\alpha(f)$ is defined as the ratio of power of the intrinsic signal in the sender over the intrinsic signal in the receiver.

Expression of the coherence based on power, interareal connectivity and the projection-source coherence

We derive the projection-source coherence based on N_p active (i.e. firing spikes) projecting neurons as follows. Let $x_i(t)$ be the activity of a single neuron in Area-1 with power spectral density $X(f)$ for all i. The cross-spectral density of the N_p projecting neurons with the signal based on all N_t neurons in Area-1 equals

$$X_{\text{proj,source}}(f) = N_p X(f) + N_p\,(N_t - 1)\, X(f) c(f)\,, \tag{28}$$

where $c(f)$ is the coherence between two individual neurons,

$$c(f) \equiv \frac{X_{i,j}(f)}{X(f)}\,. \tag{29}$$

The factor $(N_t - 1)$ accounts for the fact that each projecting neuron is fully coherent with itself. For simplicity, we assumed that the cross-spectral density between any two neurons is real-valued (i.e. all neurons are on average coherent at zero-phase). The power of the signal in the source (Area-1) equals

$$X_{\text{source,source}}(f) = N_t X(f) + N_t \left(N_t - 1\right) X(f) c(f) \,. \tag{30}$$

The power of the signal of the projection equals

$$X_{\text{proj,proj}}(f) = N_p X(f) + N_p \left(N_p - 1\right) X(f) c(f) \,. \tag{31}$$

The squared coherence now equals

$$\begin{aligned} C^2_{\text{proj,source}} &\equiv \frac{|X_{\text{proj,source}}(f)|^2}{X_{\text{proj,proj}}(f)\, X_{\text{source,source}}(f)} = \cdots \\ &\frac{\left(N_p X(f) + N_p \left(N_t - 1\right) X(f) c(f)\right)^2}{\left(N_t X(f) + N_t (N_t - 1) X(f) c(f)\right)\, \left(N_p X(f) + N_p (N_p - 1) X(f) c(f)\right)} \,, \end{aligned} \tag{32}$$

which simplifies further to

$$C^2_{\text{proj,source}} = \frac{N_p \left(1 - c(f)\right) + c(f) N_p N_t}{N_t \left(1 - c(f)\right) + c(f) N_p N_t} \,. \tag{33}$$

Plugging in $g \equiv \frac{N_p}{N_t}$ where g is the fraction of projecting neurons, we obtain,

$$C^2_{\text{proj,source}} = \frac{c(f) g (N_t - 1) + g}{c(f)(g N_t - 1) + 1} \,. \tag{34}$$

By taking the Taylor expansion around $c(f) = 0$, since the coherence between two individual neurons will be small, we obtain the first-order approximation

$$\begin{aligned} C^2_{\text{proj,source}} &\approx g + c(f)(1 - g) g N_t \\ &\approx g + g\, c(f)\, N_t \,. \end{aligned} \tag{35}$$

Here we removed the term $(1 - g)$ because we can assume that g is typically close to zero. Hence the projection-source coherence is proportional to the fraction of projecting neurons, plus the coherence times the total number of projecting neurons. We can furthermore relate $c(f)$ to the coherence of an individual neuron with the total signal in Area-1 (the spike field coherence). The squared-magnitude spike-field coherence can be expressed in terms of $c(f)$ as

$$\begin{aligned} \phi^2 &= \frac{\left((N_t - 1)\, c(f) X(f) + X(f)\right)^2}{\left(N_t X(f) + (N_t - 1)\, N_t X(f) c(f)\right) X(f)} \\ &= c(f) \left(1 - \frac{1}{N_t}\right) + \frac{1}{N_t} \,. \end{aligned} \tag{36}$$

Note that we used here

$$\begin{aligned} X_{\text{source},i} &= \sum_{j=1}^{N_t} X_{i,j} \\ &= X(f) + (N_t - 1) X(f) c(f) \end{aligned} \tag{37}$$

because we assumed all cross-spectra to be real-valued. Furthermore the total power in the source can be decomposed as

$$\begin{aligned} X_{\text{source,source}} &= \sum_{j=1}^{N_t} \sum_{i=1}^{N_t} X_{i,j} \\ &= N_t X(f) + N_t (N_t - 1) X(f) c(f) \,. \end{aligned} \tag{38}$$

Solving Eq. 36 for $c(f)$ yields

$$c(f) = \frac{\phi^2 N_t - 1}{N_t - 1}\,, \tag{39}$$

where $\phi^2 N_t \geq 1$. Plugging this into Eq. 35 we obtain the approximation

$$\begin{aligned} C^2_{\text{proj,source}} &\approx g + g\left(\phi^2 N_t - 1\right) \\ &\approx g\, N_t\, \phi^2\,. \end{aligned} \tag{40}$$

We thus obtain

$$C^2(f) \approx w^2(\alpha(f)+1)^2\phi^2\, g\, N_t\,. \tag{41}$$

Expression of the power on spike-field coherence.

We further expect $\alpha(f)$ to be proportional to ϕ^2: Let be $\phi(f)$ here is the consistency of single spikes (estimated by spike-field PPC) and divide the population into N_t spike trains of single spikes. The power due to the oscillation that is projected equals

$$\begin{aligned} S_{\text{osc}}(f) &= N_p X(f) + N_p(N_p-1)c(f)X(f) \\ &\approx N_p X(f) + N_p(N_p-1)\phi^2(f)X(f)\,. \end{aligned} \tag{42}$$

The power due to the background equals

$$\begin{aligned} S_{\text{background}}(f) &= N_p X(f) + N_p(N_p-1)c_{\text{background}}(f)X(f) \\ &\approx N_p X(f) + N_p(N_p-1)\phi^2_{\text{background}}(f)X(f)\,. \end{aligned} \tag{43}$$

Here $\phi_{\text{background}}$ is the spike-field coherence related to the background $1/f$ fluctuations, which may be non-zero. We note that if N_p is large enough, we have

$$\alpha(f) \to \frac{\phi^2(f)}{\phi^2_{\text{background}}}\,. \tag{44}$$

However, for small N_p, we obtain the first-order Taylor expansion

$$\alpha(f) \approx 1 + (N_p - 1)\left(\phi^2(f) - \phi^2_{\text{background}}\right)\,. \tag{45}$$

In this case the SOS depends on N_p. The reason for that is that when N_p is small, the contribution of the phase consistency across neurons is relatively small and the intrinsic power due to the individual energy contributions weighs in.

Non-linear dependence of coherence on spike-field coherence and connection weight.

Because the connection weight w should be proportional to the total number of projection neurons (Markov et al., 2011), we therefore expect coherence to be proportional to w and ϕ. Combining all results we obtain:

$$C^2(f) \propto w^4\phi^4\,. \tag{46}$$

The factor ϕ^4 follows from the dependence of α on ϕ^2 and $C_{\text{proj,source}}(f)$ on ϕ^2. The factor w^4 follows from the dependence of $C^2(f)$ on w^2, the dependence of $\alpha(f)$ on N_p and therefore w, and the dependence of $C_{\text{proj,source}}(f)$ on N_p and therefore w. When the number of projection neurons N_p and $\phi(f)$ is sufficiently high, the projection-source coherence $C_{\text{proj,source}}(f)$ should converge to one, and $\alpha(f)$ to $\frac{\phi^2(f)}{\phi^2_{\text{background}}}$. In that regime we obtain

$$C^2(f) \propto w^2\phi^2\,. \tag{47}$$

Linear mixture with intrinsic noise in Area-1: simulations

For the purpose of simulations, we model the signal in Area-1 as follows:

$$z_1(t) \equiv s_1(t) + \sqrt{(1-\gamma)}\eta_1(t) + \sqrt{\gamma}\epsilon(t)\,. \tag{48}$$

The intrinsic signal $z_2(t)$ of Area-2 is defined as a linear mixture of its own background fluctuations, and the input from Area-1:

$$z_2(t) = \eta_2(t) + w(s_1(t) + \eta_1(t))\,. \tag{49}$$

For the purpose of simulations, we assume that the projected oscillatory component $s_1(t)$ is fully coherent with the oscillatory process in the sending area. We now obtain

$$\begin{aligned} Z_{12}(f) &= w(S_{11}(f) + \sqrt{(1-\gamma)}H(f)) \\ &= w(\alpha(f) + \sqrt{(1-\gamma)})H(f)\,. \end{aligned} \tag{50}$$

Since $\eta_1(t)$ and $\gamma_1(t)$ are uncorrelated:

$$\begin{aligned} Z_{11}(f) &= S_{11}(f) + (1-\gamma)H(f) + \gamma H(f) && (51)\\ &= S_{11}(f) + H(f) \\ &= (1+\alpha(f))H(f)\,; && (52) \end{aligned}$$

$$\begin{aligned} Z_{22}(f) &= H(f) + w^2(S_{11}(f) + H(f)) && (53)\\ &= H(f) + w^2 Z_{11}(f) \\ &= (1 + w^2(1+\alpha))H(f)\,. \end{aligned}$$

The squared coherence $C_{12}(f)$ now simplifies as

$$C^2(f) = \frac{w^2(\alpha + \sqrt{(1-\gamma)})^2}{(1+\alpha(f))(1+w^2(1+\alpha(f)))}\,. \tag{54}$$

Plugging in $\alpha(f) = 0$ we obtain

$$C^2(f) = \frac{w^2(1-\gamma)}{(1+w^2)}\,. \tag{55}$$

This is comparable to Eq. 24.

Synthetic signals

Pink-noise signals

The background fluctuations in Figure 2 to 5 were simulated as $1/f^{2/3}$ pink-noise processes. For every trial we generate a trace of white noise sample points. Each traces was Fourier transformed. The complex coefficients of the positive frequencies were multiplied by the $1/f^{2/3}$-function. By concatenating the resulting coefficients with a flipped version of their complex conjugated we obtain a spectrum following the $1/f^{2/3}$-function. By inverse Fourier transforming the resulting spectrum we obtain a time series.

AR(2) model

The oscillatory processes in Figure 2 to 5 were simulated using an AR(2)-model. The AR(2)-model is defined as

$$x_t = a_1 x_{t-2} + a_2 x_{t-1} + \eta_t \tag{56}$$

where $\eta(t)$ is a white noise process with zero mean. To obtain a stationary signal, the roots must lie within the unit circle. If the AR process has complex conjugated roots it becomes a stochastic noise driven oscillator. The eigenvalues determine the strength of the oscillations.

Wilson-Cowan model

In this section we summarize the population model from Figure 4. The Wilson Cowan Model is a stochastic network model of nonlinear neuron models. It is often used to demonstrate the appearance of oscillations on a network scale (Powanwe and Longtin, 2019; Wallace et al., 2011; Wilson and Cowan, 1972). Each area shown in Figure 4 is modeled by a Wilson-Cowan model, composed of fully connected N_e excitatory and N_i inhibitory neurons. The neurons are modeled as two-state Markov processes (one active and one quiescent state). The transition probability of neuron i to change from the active to the quiescent state equals:

$$P_i(active \rightarrow quiescent, dt) = \alpha_i dt\,. \tag{57}$$

Whereas the transition probability of neuron i to change from the quiescent to the active state is as follows:

$$P_i(quiescent \rightarrow active, dt) = \beta_i f_i\left(s_i(t)\right) dt\,. \tag{58}$$

Here the activation function is defined:

$$f(s) \equiv \frac{1}{1+e^{-s}} \tag{59}$$

The total input current s_E to excitatory neurons and s_I to inhibitory neurons is given by:

$$s_E(t) = \frac{W_{ee}}{N_e}k(t) - \frac{W_{ei}}{N_I}l(t) + h_E \tag{60}$$

and

$$s_I(t) = \frac{W_{ie}}{N_e}k(t) - \frac{W_{ii}}{N_I}l(t) + h_I\,, \tag{61}$$

where h_I and h_E are the external input current to the correspondent neuron types. The number of active excitatory neurons is referred to as $k(t)$ and the number of active inhibitory neurons as $l(t)$. The synaptic strength from excitatory neurons to inhibitory neurons is W_{ie}, accordingly W_{ei} is the synaptic strength from inhibitory neurons to excitatory neurons. The total synaptic weight between excitatory neurons is referred to as W_{ee}, whereas the total synaptic weight between excitatory neurons is referred to as W_{ii}.

The model determines the rates of transition between states by the variables alpha and beta. However, since biological networks are stochastic processes, it is necessary to randomize the time of the next event. We achieved this by running the simulation with a Gillespie algorithm (Gillespie, 1977). In the scenario of "synaptic mixing with entrainment", the excitatory neurons from Area-1 formed connections with the excitatory neurons of Area-2. This changes equation 60 for region 2 as follows:

$$s_{E,2}(t) = \frac{W_{ee,2}}{N_{e,2}}k_2(t) - \frac{W_{ei,2}}{N_{I,2}}l_2(t) + \frac{W_{ee,1}}{N_{e,1}}k_1(t) + h_{E,2}\,. \tag{62}$$

Whereas the neurons within an area are all-to-all connected, the inter-regional connection rate in 6 B to D (right) is 5%. For simplification, each connection is represented in the LFP signal as one synapse. We calculated the LFP signal by convolving every incoming spike to an area with an alpha function $\alpha(t) \equiv g(e^{-t/\tau_1} - e^{-t/\tau_2})$. The variable t is defined as the time relative to the spike onset and $\alpha(t) = 0$ for $t < 0$. The factor g is defined as $g = -1$ for inhibitory synapses and $g = 1$ for excitatory synapses. Finally, the synaptic potentials of all input connections within an area are summed up to calculate an overall LFP signal of the corresponding area.

Each simulated area consists of 800 excitatory and 200 inhibitory neurons. The neurons within one area are fully connected. In Figure 6 A-D, each neuron in Area-2 received inputs from 40 randomly chosen excitatory neurons within Area-1. All simulations in Figure 6 have the following parameter values, $W_{ee} = 25.4$, $W_{ii} = 1.3$, $W_{ei} = 24.3$, $W_{ie} = 30$, $h_E = -3.8$, $h_I = -3.8$. Areas oscillating in the beta-frequency band have parameter values $\alpha_e = 0.038$, $\alpha_i = 0.076$, $\beta_e = 0.379$ $\beta_i = 0.758$. Areas oscillating in the gamma-frequency band have parameter values $\alpha_e = 0.1$, $\alpha_i = 0.2$, $\beta_e = 1$ $\beta_i = 2$.

Extension to spiking activity

The same model developed for field-field coherence should apply to spiking activity, if spiking relates in a linear or sigmoidal way to synaptic inputs. Consider that $z_2(t)$ represents the average voltage fluctuations in the receiver. If population spiking activity is a linear function of $z_2(t)$, i.e. $y_2(t) = z_2(t)$, then the same equation for coherence applies. Because spiking activity is stochastic and sparse for a single neuron, a population of neurons will contain additional variance that suppresses the coherence, i.e. we can write

$$y_2(t) = s_2(t) + \xi_2(t) + w\left(s_1^*(t) + \epsilon_1(t)\right) . \tag{63}$$

where $s_2(t)$ is the intrinsic signal in the receiver. This distortion ξ_2, which should decrease with the number of neurons, will decrease the interareal spike-field coherence by increasing the intrinsic power in the receiver. Next, consider the case where the population spiking activity is a standard sigmoidal activation function of $z_2(t)$, i.e.

$$y_2(t) = \sigma(z_2(t)) , \tag{64}$$

where $\sigma(x) \equiv 1/(1 + \exp(-x))$. In analogy to the data processing equality, we expect that the coherence after the transformation should always be lower than in the linear case, because the signal gets distorted by the sigmoid transformation, and coherence expresses the amount of variance that can be explained by linear prediction. Assuming that w is relatively small we can make a Taylor-expansion around $w = 0$ and obtain

$$y_2(t) = \sigma(s_2(t) + \xi_2(t)) + w\, s_1^*(t)\, \dot{\sigma}(s_2(t) + \xi_2(t)) ,$$

where $\dot{\sigma}(x)$ denotes the first derivative of the sigmoid function at x. Note that

$$\begin{aligned} &\mathrm{E}\left\{s_1^*(t) s_1^*(t+\tau)\, \dot{\sigma}(s_2(t+\tau) + \xi_2(t+\tau))\right\} \\ &= \mathrm{E}\{s_1^*(t) s_1^*(t+\tau)\}\, \mathrm{E}\{\dot{\sigma}(s_2(t+\tau) + \xi_2(t+\tau))\} . \end{aligned} \tag{65}$$

We can thus scale the signal as follows. Define a new transformation function by scaling inside the sigmoid as

$$v(x) = \sigma\left(\frac{x}{\mathrm{E}\{\dot{\sigma}(x)\}}\right) . \tag{66}$$

Assuming that $s_1(t)$ and $s_2(t)$ are statistically independent, we can see that the resulting coherence between $z_1(t)$ and $y_2(t)$ now equals

$$C^2_{\text{Area-1-LFP, Area-2-Spikes}} \approx \frac{C^2_{\text{proj,source}} w^2}{\frac{S^v_{22}(f)}{S_{11}(f)} + w^2} . \tag{67}$$

where $S^v_{22}(f)$ is the spectral density function of $v(s_2(t) + \xi_2(t))$. Here, we can recognize that the equation has the same form above and is scaled by the weight and the projection-source coherence; for small w this dependence is linear. To conclude, in case of a sigmoid input-output curve:

1. For small w, the squared spike-field coherence scales with w^2 and $C^2_{\text{proj,source}}$, similar to the case of field-field coherence.
2. Coherence between Area-1 LFP and a population of spikes in Area-2 will be lower compared to field-field coherence because the spikes will be a noisy approximation of the input signal (for a finite population). Suppose that $y_2(t)$ reflects the superposition of spiking traces from a population of neurons. The number of neurons that we superimpose in the receiver has two effects: (i) If we assume that each neuron receives the same input, then adding more neurons increases the coherence between the Area-1 LFP and the Area-2 spikes, because the population sum becomes a more accurate approximation of the LFP (Zeitler et al., 2006; Vinck et al., 2012) (i.e. in the equation above, $\xi(t)$ will decrease). (ii) Each neuron the receiver may receive synaptic inputs from only a few Area-1 projection neurons. That is, the number of projection neurons sending inputs to a single neuron in the receiver will now be equal to $N_p k$, where k is the fraction of projection neurons that goes to a single neuron in the receiver. Therefore, $C^2_{\text{proj,source}}$ may be very small for a single neuron, and it should increase with number of neurons we superimpose. Thus, the coherence between the sum of a population of Area-2 neurons and the Area-1 LFP should depend in a non-linear way on the number of neurons we sum over.

3. We expect that if the spikes are a non-linear function of the input signal, there will be a frequency-dependent distortion in interareal coherence; we expect this always to be a reduction, which remains to be proven. The distortion might be greater at higher frequencies, because the sigmoid transformation of low-frequency fluctuations can increase energy at high frequencies.

References

Abeles, M., 1982. Role of the cortical neuron: integrator or coincidence detector? Isr. Med. Assoc. J 18, 83.

Akam, T.E., Kullmann, D.M., 2012. Efficient "Communication through Coherence" Requires Oscillations Structured to Minimize Interference between Signals. PLOS Comp. Biol. 8, e1002760.

Babapoor-Farrokhran, S., Vinck, M., Womelsdorf, T., Everling, S., 2017. Theta and beta synchrony coordinate frontal eye fields and anterior cingulate cortex during sensorimotor mapping. Nature Communications , 13967.

Bastos, A.M., Vezoli, J., Bosman, C.A., Schoffelen, J.M., Oostenveld, R., Dowdall, J.R., De Weerd, P., Kennedy, H., Fries, P., 2015. Visual areas exert feedforward and feedback influences through distinct frequency channels. Neuron 85, 390–401.

Batista-Brito, R., Zagha, E., Ratliff, J.M., Vinck, M., 2018. Modulation of cortical circuits by top-down processing and arousal state in health and disease. Current opinion in neurobiology 52, 172–181.

Bonnefond, M., Kastner, S., Jensen, O., 2017. Communication between brain areas based on nested oscillations. eneuro 4.

Börgers, C., Kopell, N.J., 2008. Gamma oscillations and stimulus selection. Neural Comput 20, 383–414.

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

Bragin, A., Jandó, G., Nádasdy, Z., Hetke, J., Wise, K., Buzsáki, G., 1995. Gamma (40-100 hz) oscillation in the hippocampus of the behaving rat. J Neurosci 15, 47–60.

Bressler, S.L., 1995. Large-scale cortical networks and cognition. Brain Research Reviews 20, 288–304.

Bressler, S.L., Coppola, R., Nakamura, R., 1993. Episodic multiregional cortical coherence at multiple frequencies during visual task performance. Nature 366, 153–156.

Bressler, S.L., Richter, C.G., Chen, Y., Ding, M., 2006. Top-down cortical influences in visual expectation, in: IJCNN'06., pp. 188–194.

Brovelli, A., Ding, M., Ledberg, A., Chen, Y., Nakamura, R., Bressler, S.L., 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.

Brunet, N.M., Bosman, C.A., Vinck, M., Roberts, M., Oostenveld, R., Desimone, R., De Weerd, P., Fries, P., 2014. Stimulus repetition modulates gamma-band synchronization in primate visual cortex. Proc. Natl. Acad. Sci. U.S.A. 111, 3626–3631.

Buffalo, E.A., Fries, P., Landman, R., Buschman, T.J., Desimone, R., 2011. Laminar differences in gamma and alpha coherence in the ventral stream. Proc. Natl. Acad. Sci. U.S.A. 108, 11262–11267.

Burchell, T.R., Faulkner, H.J., Whittington, M.A., 1998. Gamma frequency oscillations gate temporally coded afferent inputs in the rat hippocampal slice. Neurosci Lett 255, 151–4.

Burns, S.P., Xing, D., Shapley, R.M., 2011. Is gamma-band activity in the local field potential of V1 cortex a "clock" or filtered noise? J. Neurosci. 31, 9658–9664.

Buschman, T.J., Miller, E.K., 2007. Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. Science 315, 1860–2.

Buzsáki, G., 2006. Rhythms of the Brain. Oxford University Press, USA.

Buzsáki, G., Anastassiou, C.A., Koch, C., 2012. The origin of extracellular fields and currents–EEG, ECoG, LFP and spikes. Nat. Rev. Neurosci. 13, 407–420.

Buzsáki, G., Draguhn, A., 2004. Neuronal oscillations in cortical networks. Science 304, 1926–1929.

Buzsáki, G., Schomburg, E.W., 2015. What does gamma coherence tell us about inter-regional neural communication? Nature neuroscience 18, 484–489.

Carmichael, J.E., Yuen, M.M., Van Der Meer, M.A., 2019. Piriform cortex provides a dominant gamma lfp oscillation in the anterior limbic system. BioRxiv , 861021.

Chalk, M., Herrero, J.L., Gieselmann, M.A., Delicato, L.S., Gotthardt, S., Thiele, A., 2010. Attention reduces stimulus-driven gamma frequency oscillations and spike field coherence in V1. Neuron 66, 114–25.

Chiu, C.Q., Lur, G., Morse, T.M., Carnevale, N.T., Ellis-Davies, G.C., Higley, M.J., 2013. Compartmentalization of gabaergic inhibition by dendritic spines. Science 340, 759–762.

Christensen, J.R., Larsen, K.B., Lisanby, S.H., Scalia, J., Arango, V., Dwork, A.J., Pakkenberg, B., 2007. Neocortical and hippocampal neuron and glial cell numbers in the rhesus monkey. The Anatomical Record: Advances in Integrative Anatomy and Evolutionary Biology: Advances in Integrative Anatomy and Evolutionary Biology 290, 330–340.

Colgin, L., Denninger, T., Fyhn, M., Hafting, T., Bonnevie, T., Jensen, O., Moser, M., Moser, E., 2009. Frequency of gamma oscillations routes flow of information in the hippocampus. Nature 462, 353–357.

Dann, B., Michaels, J.A., Schaffelhofer, S., Scherberger, H., 2016. Uniting functional network topology and oscillations in the fronto-parietal single unit network of behaving primates. Elife 5, e15719.

Donoghue, J.P., Sanes, J.N., Hatsopoulos, N.G., Gaal, G., 1998. Neural discharge and local field potential oscillations in primate motor cortex during voluntary movements. J Neurophysiol 79, 159–173.

Einevoll, G.T., Kayser, C., Logothetis, N.K., Panzeri, S., 2013. Modelling and analysis of local field potentials for studying the function of cortical circuits. Nat. Rev. Neurosci. 14, 770–785.

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

Ferro, D., van Kempen, J., Boyd, M., Panzeri, S., Thiele, A., 2020. Directed information exchange between cortical layers in macaque v1 and v4 and its modulation by selective attention. bioRxiv .

Fries, P., 2005. A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. Trends Cogn. Sci. 9, 474–480.
Fries, P., 2009. Neuronal gamma-band synchronization as a fundamental process in cortical computation. Annu. Rev. Neurosci. 32, 209–224.
Fries, P., 2015. Rhythm for Cognition: Communication Through Coherence. Neuron 88, 220–235. arXiv:15334406.
Geweke, J., 1982. Measurement of linear dependence and feedback between multiple time series. Journal of the American Statistical Association 77, 304–313.
Gillespie, D.T., 1977. Exact stochastic simulation of coupled chemical reactions. The journal of physical chemistry 81, 2340–2361.
Gray, C., McCormick, D., 1996. Chattering cells: superficial pyramidal neurons contributing to the generation of synchronous oscillations in the visual cortex. Science 274, 109.
Gray, C.M., König, P., Engel, A.K., Singer, W., 1989. Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. Nature 338, 334–337.
Gregoriou, G.G., Gotts, S.J., Zhou, H., Desimone, R., 2009. High-frequency, long-range coupling between prefrontal and visual cortex during attention. Science 324, 1207–1210.
Grothe, I., Neitzel, S.D., Mandon, S., Kreiter, A.K., 2012a. Switching neuronal inputs by differential modulations of gamma-band phase-coherence. J. Neurosci. 32, 16172–16180.
Grothe, I., Neitzel, S.D., Mandon, S., Kreiter, A.K., 2012b. Switching Neuronal Inputs by Differential Modulations of Gamma-Band Phase-Coherence. J. Neurosci. 32, 16172–16180.
Hagan, M.A., Dean, H.L., Pesaran, B., 2012. Spike-field activity in parietal area lip during coordinated reach and saccade movements. Journal of neurophysiology 107, 1275–1290.
Han, Y., Kebschull, J.M., Campbell, R.A., Cowan, D., Imhof, F., Zador, A.M., Mrsic-Flogel, T.D., 2018. The logic of single-cell projections from visual cortex. Nature 556, 51–56.
Haufe, S., Nikulin, V.V., Nolte, G., 2012. Alleviating the influence of weak data asymmetries on granger-causal analyses, in: Latent Variable Analysis and Signal Separation. Springer, pp. 25–33.
Henrie, J.A., Shapley, R., 2005. LFP Power Spectra in V1 Cortex: The Graded Effect of Stimulus Contrast. J. Neurophysiol. 94, 479–490.
Hermes, D., Miller, K., Wandell, B., Winawer, J., 2015. Stimulus Dependence of Gamma Oscillations in Human Visual Cortex. Cereb. Cortex. 25, 2951–2959.
Johnson, P.B., Ferraina, S., Bianchi, L., Caminiti, R., 1996. Cortical networks for visual reaching: physiological and anatomical organization of frontal and parietal lobe arm regions. Cerebral cortex 6, 102–119.
Katzner, S., Nauhaus, I., Benucci, A., Bonin, V., Ringach, D.L., Carandini, M., 2009. Local origin of field potentials in visual cortex. Neuron 61, 35–41.
van Kerkoerle, T., Self, M.W., Dagnino, B., Gariel-Mathis, M.A., Poort, J., van der Togt, C., Roelfsema, P.R., 2014. Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. Proc. Natl. Acad. Sci. U.S.A. , 201402773.
Kreiter, A.K., 2006. How do we model attention-dependent signal routing? Neural networks 19, 1443–1444.
Lindén, H., Tetzlaff, T., Potjans, T.C., Pettersen, K.H., Grün, S., Diesmann, M., Einevoll, G.T., 2011. Modeling the Spatial Reach of the LFP. Neuron 72, 859–872.
Livingstone, M.S., 1996. Oscillatory firing and interneuronal correlations in squirrel monkey striate cortex. J. Neurophysiol. 75, 2467–2485.
Luck, S.J., Chelazzi, L., Hillyard, S.A., Desimone, R., 1997. Neural mechanisms of spatial selective attention in areas v1, v2, and v4 of macaque visual cortex. Journal of neurophysiology 77, 24–42.
Lund, J.S., Angelucci, A., Bressloff, P.C., 2003. Anatomical substrates for functional columns in macaque monkey primary visual cortex. Cereb. Cortex. 13, 15–24.
Luppino, G., Murata, A., Govoni, P., Matelli, M., 1999. Largely segregated parietofrontal connections linking rostral intraparietal cortex (areas aip and vip) and the ventral premotor cortex (areas f5 and f4). Experimental Brain Research 128, 181–187.
Lur, G., Vinck, M.A., Tang, L., Cardin, J.A., Higley, M.J., 2016. Projection-Specific Visual Feature Encoding by Layer 5 Cortical Subnetworks. Cell Reports 14, 2538–2545.
Markov, N., Misery, P., Falchier, A., Lamy, C., Vezoli, J., Quilodran, R., Gariel, M., Giroud, P., Ercsey-Ravasz, M., Pilaz, L., et al., 2011. Weight consistency specifies regularities of macaque cortical networks. Cerebral cortex 21, 1254–1272.
Markov, N.T., Vezoli, J., Chameau, P., Falchier, A., Quilodran, R., Huissoud, C., Lamy, C., Misery, P., Giroud, P., Ullman, S., et al., 2014. Anatomy of hierarchy: Feedforward and feedback pathways in macaque visual cortex. Journal of Comparative Neurology 522, 225–259.
McGinley, M.J., Vinck, M., Reimer, J., Batista-Brito, R., Zagha, E., Cadwell, C.R., Tolias, A.S., Cardin, J.A., McCormick, D.A., 2015. Waking State: Rapid Variations Modulate Neural and Behavioral Responses. Neuron 87, 1143–1161.
Mejias, J.F., Murray, J.D., Kennedy, H., Wang, X.J., 2016. Feedforward and feedback frequency-dependent interactions in a large-scale laminar network of the primate cortex. Science advances 2, e1601335.
Michalareas, G., Vezoli, J., van Pelt, S., Schoffelen, J.M., Kennedy, H., Fries, P., 2016. Alpha-Beta and Gamma Rhythms Subserve Feedback and Feedforward Influences among Human Visual Cortical Areas. Neuron 89, 384–397. arXiv:15334406.
Miller, E.K., Wilson, M.A., 2008. All my circuits: using multiple electrodes to understand functioning neural networks. Neuron 60, 483–8.
Mitzdorf, U., 1985. Current source-density method and application in cat cerebral cortex: investigation of evoked potentials and EEG phenomena. Physiol. Rev. 65, 37–100.
Montgomery, S.M., Buzsáki, G., 2007. Gamma oscillations dynamically couple hippocampal ca3 and ca1 regions during memory task performance. Proc. Natl. Acad. Sci. U.S.A. 104, 14495–14500.
Murray, J.D., Bernacchia, A., Freedman, D.J., Romo, R., Wallis, J.D., Cai, X., Padoa-Schioppa, C., Pasternak, T., Seo, H., Lee, D., et al., 2014. A hierarchy of intrinsic timescales across primate cortex. Nature neuroscience 17, 1661.
Murthy, V.N., Fetz, E.E., 1996. Oscillatory activity in sensorimotor cortex of awake monkeys: synchronization of local field potentials and relation to behavior. J. Neurophysiol. 76, 3949–3967.
Nicolelis, M.A., Baccala, L.A., Lin, R., Chapin, J.K., 1995. Sensorimotor encoding by synchronous neural ensemble activity at multiple levels of the somatosensory system. Science 268, 1353–1358.
Nolte, G., Bai, O., Wheaton, L., Mari, Z., Vorbach, S., Hallett, M., 2004. Identifying true brain interaction from EEG data using the imaginary part

of coherency. Clin Neurophysiol 115, 2292–2307.
Nunez, P.L., Srinivasan, R., 2006. Electric fields of the brain: the neurophysics of EEG? Oxford University Press.
Olcese, U., Bos, J.J., Vinck, M., Lankelma, J.V., van Mourik-Donga, L.B., Schlumm, F., Pennartz, C.M., 2016. Spike-based functional connectivity in cerebral cortex and hippocampus: loss of global connectivity is coupled to preservation of local connectivity during non-rem sleep. Journal of Neuroscience 36, 7676–7692.
Onorato, I., Neuenschwander, S., Hoy, J., Lima, B., Rocha, K.S., Broggini, A.C., Uran, C., Spyropoulos, G., Klon-Lipok, J., Womelsdorf, T., Fries, P., Niell, C., Singer, W., Vinck, M., 2020. A distinct class of bursting neurons with strong gamma synchronization and stimulus selectivity in monkey v1. Neuron 105, 180–197.
Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.M., 2011. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci 2011, 156869.
Palmigiano, A., Geisel, T., Wolf, F., Battaglia, D., 2017. Flexible information routing by transient synchrony. Nature neuroscience 20, 1014.
Parabucki, A., Lampl, I., 2017. Volume conduction coupling of whisker-evoked cortical lfp in the mouse olfactory bulb. Cell reports 21, 919–925.
Pesaran, B., Vinck, M., Einevoll, G., Sirota, A., Fries, P., Siegel, M., Truccolo, W., Schroeder, C., Srinivasan, R., 2018. Investigating large-scale brain dynamics using field potential recordings: analysis and interpretation. Nat. Neurosci. .
Peter, A., Uran, C., Klon-Lipok, J., Roese, R., Van Stijn, S., Barnes, W., Dowdall, J.R., Singer, W., Fries, P., Vinck, M., 2019. Surface color and predictability determine contextual modulation of v1 firing and gamma oscillations. eLife 8, e42101.
Phillips, J.M., Vinck, M., Everling, S., Womelsdorf, T., 2014. A long-range fronto-parietal 5-to 10-hz network predicts "top-down" controlled guidance in a task-switch paradigm. Cerebral Cortex 24, 1996–2008.
Pike, F., Goddard, R., Suckling, J., Ganter, P., Kasthuri, N., Paulsen, O., 2000. Distinct frequency preferences of different types of rat hippocampal neurones in response to oscillatory input currents. The Journal of Physiology 529, 205.
Powanwe, A.S., Longtin, A., 2019. Determinants of brain rhythm burst statistics. Scientific Reports 9, 1–23.
Ray, S., Maunsell, J.H., 2015. Do gamma oscillations play a role in cerebral cortex? Trends in cognitive sciences 19, 78–85.
Richter, C.G., Coppola, R., Bressler, S.L., 2018. Top-down beta oscillatory signaling conveys behavioral context in early visual cortex. Sci. Rep. 8, 6991.
Salazar, R., Dotson, N., Bressler, S., Gray, C., 2012a. Content-specific fronto-parietal synchronization during visual working memory. Science 338, 1097–1100.
Salazar, R.F., Dotson, N.M., Bressler, S.L., Gray, C.M., 2012b. Content-Specific Fronto-Parietal Synchronization During Visual Working Memory. Science 338, 1097–1100.
Saleem, A.B., Lien, A.D., Krumin, M., Haider, B., Rosón, M.R., Ayaz, A., Reinhold, K., Busse, L., Carandini, M., Harris, K.D., 2017. Subcortical Source and Modulation of the Narrowband Gamma Oscillation in Mouse Visual Cortex. Neuron 93, 315–322.
Salinas, E., Sejnowski, T.J., 2001. Correlated neuronal activity and the flow of neural information. Nat Rev Neurosci 2, 539–50.
Scherberger, H., Jarvis, M.R., Andersen, R.A., 2005. Cortical local field potential encodes movement intentions in the posterior parietal cortex. Neuron 46, 347–354.
Schomburg, E.W., Fernández-Ruiz, A., Mizuseki, K., Berényi, A., Anastassiou, C.A., Koch, C., Buzsáki, G., 2014. Theta phase segregation of input-specific gamma patterns in entorhinal-hippocampal networks. Neuron 84, 470–485.
Siegle, J.H., Jia, X., Durand, S., Gale, S., Bennett, C., Graddis, N., Heller, G., Ramirez, T.K., Choi, H., Luviano, J.A., et al., 2019. A survey of spiking activity reveals a functional hierarchy of mouse corticothalamic visual areas. bioRxiv , 805010.
Singer, W., Gray, C.M., 1995. Visual feature integration and the temporal correlation hypothesis. Annu. Rev. Neurosci. 18, 555–586.
Sirota, A., Montgomery, S., Fujisawa, S., Isomura, Y., Zugaro, M., Buzsáki, G., 2008a. Entrainment of neocortical neurons and gamma oscillations by the hippocampal theta rhythm. Neuron 60, 683–697.
Sirota, A., Montgomery, S., Fujisawa, S., Isomura, Y., Zugaro, M., Buzsáki, G., 2008b. Entrainment of Neocortical Neurons and Gamma Oscillations by the Hippocampal Theta Rhythm. Neuron 60, 683–697.
Spyropoulos, G., Dowdall, J.R., Schölvinck, M.L., Bosman, C.A., Lima, B., Peter, A., Onorato, I., Klon-Lipok, J., Roese, R., Neuenschwander, S., Wolf, S., Vinck, M., Fries, P., 2020. Spontaneous variability in gamma dynamics described by a linear harmonic oscillator driven by noise. bioRxiv , 793729.
Steriade, M., McCormick, D.A., Sejnowski, T.J., 1993. Thalamocortical oscillations in the sleeping and aroused brain. Science 262, 679–685.
Tiesinga, P., Sejnowski, T., 2010. Mechanisms for phase shifting in cortical networks and their role in communication through coherence. Front. Hum. Neurosci 4.
Trongnetrpunya, A., Nandi, B., Kang, D., Kocsis, B., Schroeder, C.E., Ding, M., 2016. Assessing granger causality in electrophysiological data: removing the adverse effects of common signals via bipolar derivations. Frontiers in systems neuroscience 9, 189.
Varela, F., Lachaux, J.P., Rodriguez, E., Martinerie, J., 2001. The brainweb: phase synchronization and large-scale integration. Nat. Rev. Neurosci. 2, 229–239.
Vinck, M., Battaglia, F.P., Womelsdorf, T., Pennartz, C., 2012. Improved measures of phase-coupling between spikes and the Local Field Potential. J Comput Neurosci 33, 53–75.
Vinck, M., Bos, J.J., Mourik-Donga, V., Laura, A., Oplaat, K.T., Klein, G.A., Jackson, J.C., Gentet, L.J., Pennartz, C., 2016. Cell-type and state-dependent synchronization among rodent somatosensory, visual, perirhinal cortex, and hippocampus ca1. Frontiers in systems neuroscience 9, 187.
Vinck, M., Bosman, C.A., 2016. More gamma more predictions: Gamma-synchronization as a key mechanism for efficient integration of classical receptive field inputs with surround predictions. Front Syst Neurosci 10, 35.
Vinck, M., Huurdeman, L., Bosman, C.A., Fries, P., Battaglia, F.P., Pennartz, C.M., Tiesinga, P.H., 2015. How to detect the granger-causal flow direction in the presence of additive noise? Neuroimage 108, 301–318.
Vinck, M., Oostenveld, R., van Wingerden, M., Battaglia, F., Pennartz, C.M., 2011. An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias. Neuroimage 55, 1548–1565.
Vinck, M., van Wingerden, M., Womelsdorf, T., Fries, P., Pennartz, C.M., 2010. The pairwise phase consistency: a bias-free measure of rhythmic neuronal synchronization. Neuroimage 51, 112–122.

Vinck, M., Womelsdorf, T., Buffalo, E.A., Desimone, R., Fries, P., 2013a. Attentional modulation of cell-class-specific gamma-band synchronization in awake monkey area V4. Neuron 80, 1077–1089.

Vinck, M., Womelsdorf, T., Fries, P., 2013b. Gamma-band synchronization and information transmission, in: Quiroga-Quian, R., Panzeri, S. (Eds.), Principles of Neural Coding. CRC Press.

Volgushev, M., Chistiakova, M., Singer, W., 1998. Modification of discharge patterns of neocortical neurons by induced oscillations of the membrane potential. Neuroscience 83, 15–25.

Von Stein, A., Sarnthein, J., 2000. Different frequencies for different scales of cortical integration: from local gamma to long range alpha/theta synchronization. International Journal of Psychophysiology 38, 301–313.

Wallace, E., Benayoun, M., Van Drongelen, W., Cowan, J.D., 2011. Emergent oscillations in networks of stochastic spiking neurons. Plos one 6.

Wang, X.J., 2002. Pacemaker neurons for the theta rhythm and their synchronization in the septohippocampal reciprocal loop. Journal of neurophysiology 87, 889–900.

Wilson, H.R., Cowan, J.D., 1972. Excitatory and inhibitory interactions in localized populations of model neurons. Biophys. J. 12, 1–24.

Xing, D., Yeh, C.I., Burns, S., Shapley, R.M., 2012. Laminar analysis of visually evoked activity in the primary visual cortex. Proc. Natl. Acad. Sci. U.S.A. 109, 13871–13876.

Zeitler, M., Fries, P., Gielen, S., 2006. Assessing neuronal coherence with single-unit, multi-unit, and local field potentials. Neural Comput 18, 2256–2281.

6 Appendix B : supplementary tables

Tables listed in this appendix serve as supplementary material and are referenced in the General Introduction section of this thesis.

Table 1: List of some studies demonstrating oscillatory synchronization between spikes and LFPs during movement behaviors.

Subject and task details	**Recording details**	**Modulations in oscillatory synchrony**	**Interpretation**
(Murthy and Fetz, 1996) 3 macaques; reaching movements to retrieve food pieces	Sequential recordings from 134 unilateral and 42 bilateral pairs from sensorimotor cortices using tungsten electrodes	Activity of many task-related as well as task-unrelated neurons found to be synchronized to LFP oscillations (20-40 Hz)	Synchronization reflects mechanisms related to arousal/attention
(Scherberger, Jarvis and Andersen, 2005) 2 macaques; delayed reaching and saccade task	Sequential recordings from 137 sites in the parietal reach region using tungsten electrodes	Enhanced population-level spike-field beta (20-40 Hz) and reduced low (0-10 Hz) frequency coherence during planning epoch for reaches compared to saccades	Movement intention might be encoded in spike-field beta synchrony
(Pesaran, Nelson and Andersen, 2008) 2 macaques; free and instructed search tasks with a reach component	Sequential recordings from dorsal premotor area and parietal reach region using Pt/Ir elec- trodes (314 PMd spike–PRR field and 187 PRR spike–PMd field)	Increased SFC (15 Hz) during decision making task than during simple center-out reach task; Higher spike-field coherence when monkeys made free-choices than when they followed instructions	Decision making requires exchange of information between sub-populations of frontal and parietal neurons, which might be reflected by synchrony

(Wong *et al.*, 2016) 2 macaques trained to perform multiple tasks involving reach and saccade components	Recordings from 4 tungsten electrodes per session from the medial and the lateral banks of the intraparietal sulcus in the Posterior Parietal Cortex (144 unique spike-LFP pairs)	Three distant categories of neurons identified based on whether they were beta (15-25 Hz) coherent with LFPs recorded on the same bank, on the other bank, or both (dual coherent); The firing rate of dual coherent neurons predicted movement choices faster than the other neurons	Interactions of dual coherent neurons important in the decision circuitry; Coherence between the reach (medial bank) and saccade (lateral bank) systems associated with choice-predictive firing supports models of interacting selection

Table 2: List of some studies demonstrating oscillatory synchronization between spikes and LFPs during visual tasks. Some of these studies also evaluated spike-spike interactions.

Subject and task details	**Recording details**	**Modulations in oscillatory synchrony**	**Interpretation**
(Gray and Singer, 1989) 15 adult cats anaesthetized and presented visual stimuli	Recordings from thalamus and areas 17 and 18 of visual cortex (or V1 and V2) using one to five Teflon-coated platinum-iridium electrodes	Firing of neurons in the visual cortex found to be oscillatory in the gamma band (~25 to 65 Hz) and coupled to the phase and amplitude of LFP signals; Neurons recorded from thalamus exhibited non-oscillatory firing patterns	Stimulus specific intra-cortical oscillations as a mechanism to temporally coordinate activity from spatially distributed regions
(Fries, 2001) 2 monkeys; a selective visual attention task	Simultaneous recordings from four extracellular electrodes in visual area V4	Enhanced gamma (35-90 Hz) band synchronization with a simultaneous reduction in low-frequency (< 17 Hz) synchronization in neurons activated by attended stimulus compared to neurons activated by distractors in the same area	Changes in synchronization might amplify behaviorally relevant signals
(Womelsdorf *et al.*, 2006) 2 monkeys; change detection of visual stimuli	Simultaneous recordings from 4-8 electrodes in V4 (multi-units from 61 and LFPs from 64 sites, both monkeys)	Higher spike-field gamma (40 – 70 Hz) synchronization before and after change in relevant stimulus in neurons activated by attended stimulus related to shorter neuronal response latency and faster reaction times	Gamma synchrony might reflect early neuronal correlate of visuo-motor integration

(Gregoriou *et al.*, 2009) 2 macaques; selective attention task	Four tungsten microelectrodes each in frontal eye field (FEF) and V4 per session. (448 FEF spike–V4 field and 647 V4 spike–FEF field)	Attention enhanced oscillatory coupling in the gamma band (40-60 Hz), whereas decreased oscillatory coupling in the low frequency range (5-20 Hz) between areas FEF and V4 when the stimulus was presented in their joint receptive field	FEF might be a source of the attentional effects on gamma synchrony in V4 and ventral visual areas; Oscillatory synchrony might be regulating communication across brain structures
(Vinck *et al.*, 2013) 2 monkeys trained to perform selective visual attention task	Recordings from 4-8 tungsten electrodes advanced separately in V4 to analyze	Cell-type as well as task epoch specific changes in gamma synchronization (30–70 Hz); Firing rate effect observed: attention increased gamma locking in strongly activated cells and vice-versa for weakly activated cells	Cell-type specific dynamics of gamma cycle in V4 might provide insights into potential mechanisms underlying gamma synchrony
(Onorato *et al.*, 2020) 3 different animal models (2 macaques, one capuchin monkey, mice); fixation task	Recordings made from V1 using 2 to 10 microelectrodes in monkeys; two-shank 32-channel silicon probes in mice	A distinct neuron class identified based on waveform shape and burst firing propensity in V1 of monkeys and not in mice; These neurons exhibited early onset as well as stronger gamma synchrony (30–80 Hz) and had higher stimulus selectivity	This new class of excitatory neurons (putative chattering cells) in interaction with inhibitory inter-neurons might play a key role in generating strong gamma phase-locking

Table 3: List of some studies demonstrating oscillatory synchronization between spikes and LFPs during memory and other cognitive tasks. Some of these studies also evaluated spike-spike interactions.

Subject and task details	**Recording details**	**Modulations in oscillatory synchrony**	**Interpretation**
(Siapas, Lubenov and Wilson, 2005) 4 rats performing freely moving behavior in familiar environments	Recordings made using 18 individual tetrodes from medial pre-frontal cortex and hippocampus	Neurons in the medial pre-frontal cortex identified to be phase-locked to the hippocampal theta rhythm of LFPs with a delay of ~50ms, whereas they correlated with the hippocampal neurons with a delays < 150ms	Delayed correlations suggest that direct hippocampal input is important for prefrontal theta phase-locking; Theta synchronization might facilitate selective information flow and plastic changes in the circuitry for information storage
(Sirota *et al.*, 2008) Multiple experiments conducted across 28 rats and 11 mice	Multiple types of electrodes used to record from the hippocampus, pre-frontal, parietal and primary somatosensory cortex	Different groups of neocortical neurons oscillating at different frequencies in the gamma band phase-locked with the hippocampal theta rhythm	Timing of the neocortical gamma oscillators might be coordinated by the hippocampal theta and their information might be synchronously transferred to the hippocampus
(Siegel, Warden and Miller, 2009) Two macaque monkeys trained on two object visual short term memory task	Recording from 140 sites in pre-frontal cortex across sessions using arrays of 8 independently movable tungsten microelectrodes	During memory epoch, neurons were synchronized at 32 and 3 Hz; Spikes occurring during the earlier (later) phase of the 32 Hz oscillation carried more information on the identity of the object presented first (second)	Oscillatory synchronization facilitates phase-dependent encoding of objects in short term memory

7 Curriculum Vitae

Swathi Sheshadri

DOB: 19th Jan 1990,
Affiliation: Neurobiology Lab,
German Primate Center,
Cell:+4917662223384
Email: swathishesh@gmail.com
Github: SwathiSheshadri

Education

Doctoral Candidate, Theoretical and Computational Neuroscience, Georg-August University, Goettingen, 2016-present

- Dissertation: Neuron-level dynamics of oscillatory network structure and markerless tracking of kinematics during grasping
- Dissertation Advisors: Hansjoerg Scherberger, Alexander Gail and Michael Wibral

M.Sc., Electrical and Computer Engineering, National University of Singapore, Singapore, 2013-2015

- Courses include: Pattern recognition, Neural networks, Stochastic processes
- Cumulative Grade Point Average: 3.9 out of 5.0 with 5.0 being the best
- Project: Neural decoding of peripheral signals
- Project Advisor: Shih-Cheng Yen

B.E, Instrumentation Technology, Visvesvaraya Technological University, India, 2007-2011

- Courses include: Digital signal processing, Biomedical instrumentation, Medical imaging
- Cumulative Grade Point Average: 9.5 out of 10.0 with 10.0 being the best

Publications

Sheshadri S*, Dann B* and Scherberger H, (**prepared for submission**). Distinct beta and low frequency oscillatory neurons participate in a behavior dependent reconfiguration of the fronto-parietal grasping network in macaques.

Sheshadri S*, Hueser T*, Dann B*, and Scherberger H, (**in prep**). Precise markerless tracking of the hand during object grasping.

Sheshadri S, Dann B, Hueser T and Scherberger H, 2020. 3D reconstruction toolbox for behavior tracked with multiple cameras. Journal of Open Source Software, 5(45), 1849, https://doi.org/10.21105/joss.01849

Lee S, **Sheshadri S,** Xiang Z, Delgado-Martinez I, Xue N, Sun T, Thakor N.V, Yen S.C and Lee C, 2017. Selective stimulation and neural recording on peripheral nerves using flexible split ring electrodes. *Sensors and Actuators B: Chemical*, *242*, pp.1165-1170, https://doi.org/10.1016/j.snb.2016.09.127

Xiang Z, Yen S.C, **Sheshadri S,** Wang J, Lee S, Liu Y.H, Liao L.D, Thakor N.V and Lee C, 2016. Progress of Flexible Electronics in Neural Interfacing–A Self-Adaptive Non-Invasive Neural Ribbon Electrode for Small Nerves Recording. *Advanced Materials*, *28*(22), pp.4472-4479, https://doi.org/10.1002/adma.201503423

Sheshadri S, Kortelainen J, Rigosa J, Cutrone A, Bossi S, Libedinsky C, Lahiri A, Chan L, Chng K, Thakor N.V and Delgado-Martínez I, 2015, April. Classification of phases of hand grasp task by the extraction of miniature compound nerve action potentials (mCNAPs). In *2015 7th International IEEE/EMBS Conference on Neural Engineering (NER)* (pp. 593-596), IEEE, 10.1109/NER.2015.7146692

Sheshadri S, Kortelainen J, Nag S, Ng K.A, Bazley F.A, Michoud F, Patil A, Orellana J, Libedinsky C, Lahiri A and Chan L, 2014, August. Correlation between muscular and nerve signals responsible for hand grasping in non-human primates. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 2314-2317), IEEE, 10.1109/EMBC.2014.6944083

Full list: https://scholar.google.com/citations?user=1BB50RkAAAAJ&hl=en&oi=ao

Talks

Separate groups of beta and low frequency neurons form behavior- and state-dependent networks in the macaque fronto-parietal grasping circuit, at the Mainhatten Lectures in Neuroscience and whatnot, April 28, 2020 (virtual lecture)

Behavior dependent states in segregated networks of low and beta synchronized neurons in the macaque fronto-parietal grasping circuit, at the 49th Annual Meeting of the Society for Neuroscience, Chicago, October 19-23, 2019 (travel grant awarded)

Behavioral and temporal dynamics of the fronto-parietal grasping network, at the 11th Annual Primate Neurobiology Meeting, Tuebingen, March 13-14, 2018

Comparison of methods of spike-field coherence in the front-parietal grasp network, at the Ernst Struengmann Institute (ESI) for Neuroscience, Frankfurt, May 23, 2016 (in fulfilment of requirements of Graduate Training in Primate Neurobiology program)

Classification of phases of hand grasp task by the extraction of miniature compound nerve action potentials (mCNAPs), at the 7th International IEEE EMBS Conference on Neural Engineering, France, April 22-24, 2015

Posters

Marker-free primate hand tracking during reach and grasp movements of diverse objects, Bernstein Conference, Berlin, 2018 (travel grant awarded)

Behavior dependent antagonistic synchronization of beta and low frequency hub neurons in the macaque fronto-parietal grasping network, 47th Annual Meeting of the Society for Neuroscience, Washington DC, 2017 (travel grant awarded)

Condition-specific functional connectivity of the fronto-parietal grasping network, 10th Annual Primate Neurobiology Meeting, Goettingen, 2017

Behavior-dependent functional connectivity of beta and low-frequency hub units in the macaque fronto-parietal grasp network, Bernstein Conference, Goettingen, 2017

Condition-specific functional connectivity of the fronto-parietal grasping network, Primate Neurobiology Meeting, Goettingen, 2017

Decoding of motor information in non-human primates using a chronic implantable system, Symposium on Grand Challenges in Neural Technology, Singapore, 2013

Teaching Experience

Teaching Assistant

- Scherberger lab rotation (Lecture and Programming lab), 2019
- Course Instructor: Benjamin Dann

Teaching Assistant

- Motor systems (Lecture and Seminar), 2017
- Course Instructor: Hansjoerg Scherberger

Student Supervision

Master Thesis

- Viktorija Schek, 2020
  Performance optimization of a markerless 3D tracking system for grasping tasks with Object Interactions

Lab rotation

- Selene Lickfett, 2019
  Coding and decoding of hand grasping movements
- Viktorija Schek, 2018
  Evaluation of a markerless tracking system for grasp tracking with three-camera system
- Michael Lutz, 2016
  Creation of a graphical user interface for visual control of neuronal data quality

Work Experience

Internship, German Primate Center, 2015-2016
Comparison of spike-field coherence methods, learnt about non-human primate training techniques

Research Assistant, Singapore Institute of Neurotechnology, 2014-2015
Electroneurographic (ENG) and electromyographic (EMG) data acquisition and analysis from non-human primates and rodents including signal denoising, characterization and classification

Business Technology Analyst, Deloitte Consulting, 2011-2013
Technology consulting with advanced business application programming

Other Relevant Specializations and Skills

Deep learning Coursera specialization: Topics studied include hyperparameter tuning, regularization and optimization in neural networks, convolutional neural networks, sequence models
Programming ability: Matlab and Python
Language ability: Fluent in English, Intermediate (B1) in German
Organizing member: DPZ colloquium (2017-present), Women Career Networking Symposium, 2019, Goettingen