Statistical analysis of travelling waves in the monkey primary motor cortex

Jessica Chemali
Machine Learning Department
Carnegie Mellon University
Pittsburgh, PA 15213
jchemali@cs.cmu.edu

Abstract

Traveling waves are conjectured to play a computational role in information transfer within and across regions of the brain, however both their function and morphology are still poorly understood - especially in the motor cortex. Linear waves of local field potential (LFP) over beta oscillation range (15-30 Hz) are thought to be associated with movement planning when sweeping across a microelectrode array. However, it is not known whether more localized dynamics are present, and whether the different parts of the primary motor cortex spanned by the array contain different pieces of information about the upcoming movement. Previous studies that analyzed linear waves in the brain suffered from two main statistical weaknesses. First, they did not clarify the null hypothesis they were testing against and second they did not control the overall error rates of their procedures. Our first contribution is to develop a statistical procedure for detecting such waves. We then analyze the local topography of traveling waves in intracortical microelectrode recordings of the primary motor cortex when a monkey is moving a cursor to instructed directions. We do so by dividing the grid array into smaller square grids or "patches" and fitting a series of linear regression models to the phase spatio-temporal series extracted from the LFP. Our results suggest that there is indeed spatial heterogeneity across patches, both in terms of number of patterns identified per patch and their directions of propagation. This finding are a first step towards analyzing the morphology of waves in the primary cortex, and encourage future studies to investigate the information content of its different parts.

1 Introduction

Travelling waves are spatio-temporal patterns that arise naturally in oscillating media. In neuroscience, they are conjectured to play a computational role in information transfer within and across regions of the brain [1, 2]. However, even though their presence has been repeatedly detected in the visual [3], auditory [4], and motor cortices [2], as well as in subcortical structures [5], both their function to behavior and their morphology are poorly understood, especially in the motor cortex [6, 7].

Neuroscientifc methods for detecting and analyzing waves are currently under-developed compared to other fields, such as oceanography and seismology, where traveling waves are also important [8, 9, 10, 11]. Three main factors complicate their analysis in the brain: First, brain waves show transient dynamics, more so than ocean waves for example, which raises the temporal and spatial resolution requirements of recording tools and usually precludes the ability to conduct classical frequency-domain analysis. Second, capturing the correct spatial range and/or resolution of the wave of interet might be infeasible sometimes due to constraints in sensors placements, which in turn might render its mode of propagation unidentifiable. For instance, while in the visual brain

regions many types of waves have been observed (e.g. concentric [12], spiral wave [13], etc.), only linear waves have been analyzed in the motor cortex, in part because the recorded areas were too small to cover the global activity and reveal diverse wave patterns with the straightforward analysis methods. Finally, the third challenge is the high level of intrinsic noise of the underlying brain signals which makes it harder to identify the signal to analyze.

One analysis approach, common in neuroscience, is to track the temporal phase dynamics of recorded trials [14, 15]. In particular, [2] demonstrated the presence of linear traveling waves in the monkey primary motor cortex by tracking the degree of alignment of the phase gradients across the recording electrodes. Similarly, [16] demonstrated the presence of linear traveling waves within a stratum of the hippocampus by fitting a series of linear regression models to the phase spatio-temporal series and tracking their goodness of fit measures. Although crucial for interpreting the results, a complete statistical analysis was missing from both these studies. In [2] the authors chose an arbitrarily threshold above which the gradients' alignment is taken to be strong *enough*, and similarly in [16] the authors report low p-values and a high R^2 to indicate a *sufficiently* good fit. However, these methods suffer from two main statistical weaknesses. First, the statistics used to determine the presence of a wave are not absolute quantities and the strength of the patterns they indicate cannot be judged in an absolute manner but only relative to a baseline. Formally, this is equivalent to a hypothesis testing problem where the null hypothesis was unclear. The second weakness of the analyses is that they didn't account for the large number of hypothesis tests that are conducted. In both cases, the presence of a wave was judged at every time point, and given the large number of time points, multiple comparison statistical procedures should have been used to control for the overall error rates.

In this work, we build on previous literature in two main ways. First, we improve the methodology by designing a statistical procedure based on the permutation testing framework that selects the time segments where linear phase patterns are statistically significant. As we show below, this task is subtle and its success depends on, among other things, the number of electrodes chosen in the analysis, the "linear" nature of the phase data, and the cross-correlation strength between electrodes. Second, we investigate the local morphology of the wave activity in the primary motor cortex by re-analyzing the phase spatio-temporal series of local field potential (LFP) recordings from [2]. Specifically, we simplify the problem of determining the shape of the wave (e.g, concentric or spiral) by only seeking to determine whether the wave activity is heterogeneous across the array. If it is, then it would give an indication that the wave activity in the motor cortex is not limited to linear waves and it would open the door to more investigations in that direction. We do so by dividing the recording array into smaller non-overlapping square patches, and comparing the distributions of estimated wave properties both across patches and across experimental conditions. This approach allows us to analyze the LFP recordings beyond the limiting assumption that a single linear wave is present at any time point while still using simple linear estimation methods of [2, 16].

In the rest of this document, we first introduce our dateset (section 2) and carefully revisit the phase processing pipeline which consists of the extraction and unwrapping of the phase (section 3.1). We then analyze the travelling waves by fitting a series of linear regression models to the phase spatio-temporal series and provide a comparison with the original method used for this dataset [2] (section 3.2). Finally, we describe the statistical procedure used to detect the time segments with statistically significant linear wave patterns (section 5). In our results section (section 6) we show how the distribution of the test statistics under the null hypothesis can guide us to determine the patch size, and the method that balance the locality of the analysis with its power to detect meaningful patterns. Our results demonstrate that the wave activity across the array is heterogeneous and that it is generally consistent in every patch across experimental conditions.

2 Dataset description

Intracortical microelectrode measurements, organized in a 10×10 square array with 400 μm inter-electrode spacing, were recorded from the arm area of a monkey's primary motor cortex while he was performing an *instructed-delay center-out* task. This task consisted of three periods, referred to as the *hold*, *instruction* and *movement* periods. During the *hold* period, the monkey was trained to hold the cursor on a center target and wait 500 ms for the instruction cue. During the *instruction* period, the monkey was presented with one of eight evenly spaced peripheral targets and continued to

hold at the center for an additional 1000-1500 ms. At that time, the go cue was presented, signaling the monkey to begin *movement* to the peripheral target. The end of movement is also timestamped and corresponds to the time when the monkey reached the target. For the rest of the document we will refer to the set of successful trials with different target movement directions as different experimental conditions. Trials where the monkey did not successfully reach for the correct target direction were ignored.

The LFP recordings consist of 96 channels sampled at 1KHz. There are around 50 independent trials per experimental condition of around 2 seconds each. Fig. 1 shows the spectrogram averaged over all trials and all electrodes, centered around the instruction cue (time 0). Individual spectrograms (i.e, for each trial-electrode) are computed using the multitaper method with time-bandwidth product equal to 2 and with 3 tapers. Notice the peak in power between 10 and 25 Hz just after the instruction cue. This increase has been reported consistently in studies of the primary motor cortex and is associated with the arrival of upcoming movement information. While previous studies have mainly focused on this period, in this work we analyze the segments centered around the instruction cue, with 500 ms of data prior to it and 1000 ms after it. We chose this larger time span in order to investigate more globally the distribution of waves onset.

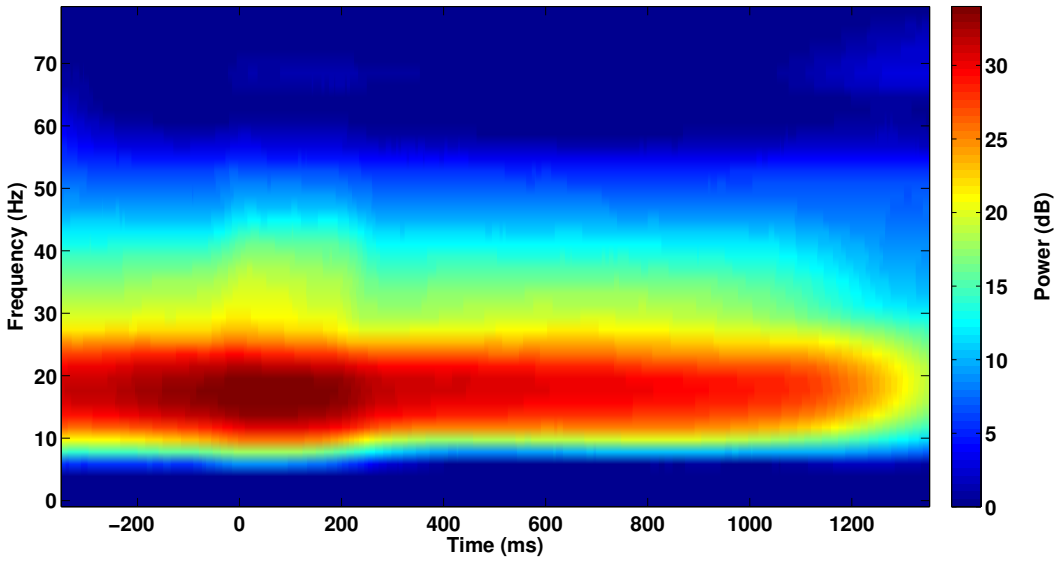

Figure 1: The spectrogram averaged over all trials and channels of the lfp dataset. Note the peak in power in the 10-25 Hz frequency band, in particular right after time 0, after the instruction cue onset.

2.1 Wave activity

[2] showed that the increase in beta power was temporarily associated with linear waves, characterized by phases that were organized linearly across the array.

Fundamentally, a wave is an orchestrated oscillatory activity that travels through a medium from one location to another. It is characterized by three quantities: Its *amplitude* defined by the distance from rest to crest, its *wavenumber* k defined to be the number of cycles per space unit (or equivalently *wavelength* $\lambda = \frac{2\pi}{k}$), and its *frequency* f defined to be the number of cycles per time unit (or equivalently *period* $T = \frac{1}{f}$). Additionally, its *speed* can be derived as $\frac{\lambda}{T} = \frac{\omega}{k}$ and is defined as the distance traveled by a given point on the wave (such as crest) per time unit.

A 2-dimensional surface representing a single linear traveling wave can be modeled by:

$$v(x, y, t) = V(k_x x + k_y y + \omega t + \phi_0), \tag{1}$$

where (x, y) is the coordinate vector, t is the time index, $\vec{k} = (k_x, k_y)$ is the wavenumber vector and ϕ_0 is the phase offset. Therefore to fully characterize the dynamics of a linear wave it suffices

to determine its wavenumber vector and its angular frequency. In Fig. 2 (top row) we illustrate a simulated linear wave. A complete cycle of the wave can be seen spanning the whole array and its peak (red part) can be seen moving as time passes from $t1$ to $t5$. The second row of the figure shows the phase of the wave. It is organized as a plane over the array and its slope remains constant over time. The third row of Fig. 2 is a compass plot of the directions in which the phase gradients at the different electrodes are pointing. For this perfect linear wave, they all coincide at 210 degrees and remain constant across time. The slope of the plane is the wave's wavenumber and in section 3.2 we show how to estimate this quantity using either linear regression or using the average spatial gradient across the array.

In this work we assume that linear waves are a reasonable approximation to segments of activity in our data, allbeit corrupted with noise and transient. The previous analysis in [2] estimated the average wavelength of waves oscillating at around 20 Hz to be approximately 10 mm, which is equivalent to an approximate speed of 0.2 mm/ms. This means that the wavelength is around twice the width of the array, and only half of the wave cycle can be seen in a single snapshot.

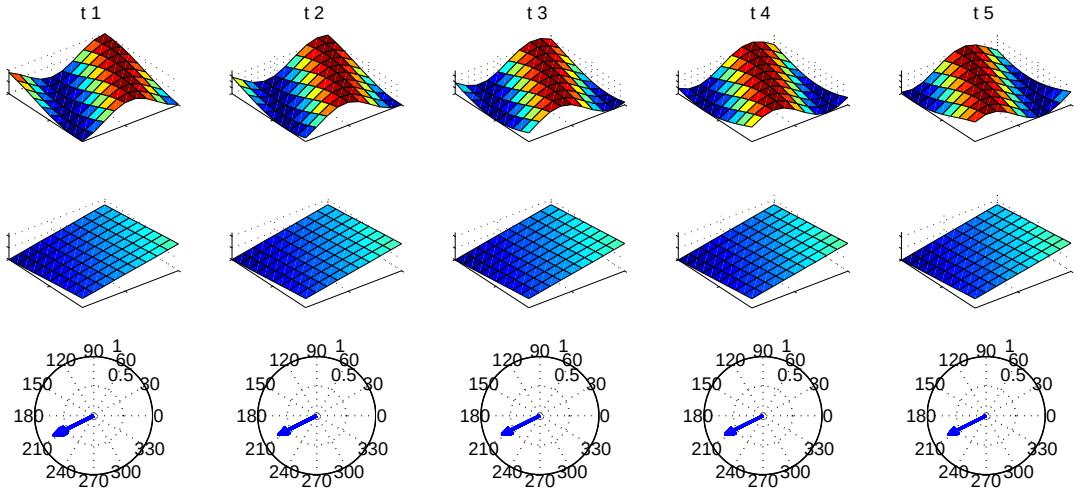

Figure 2: Simulated linear wave. **Top panel**: shows 5 time snapshots of the simulated voltage on a 10×10 array, **Middle panel**: shows the associated absolute phase which is organized as a plane of constant slope, **Bottom panel**: show the directions of the phase gradient vectors of all electrodes. All of them coincinde at around 210 degrees.

3 Background Methods

As described in section 2, linear wave dynamics are summarized by the wave's wavenumber vector $\vec{k}$ and its temporal frequency ω. Below, we describe how to estimate these quantities using 2 methods: by fitting linear regression models to the phase data as in [16, 17] or by numerically computing the average spatial phase gradient as in [2]. In this section, we first describe the phase processing required to extract the phase and make it adequate for use in estimating the wave properties using the above mentioned methods. We then describe our statistical procedure that allow to select the statistically significant linear wave segments in our dataset.

3.1 Phase extraction and processing

3.1.1 Phase extraction

Let $v(x, y, t)$ denote the LFP signal, where (x, y) denote the position of the electrode on the 2-dimensional recording grid and t denotes the discrete time index.

Given the property of real-valued signals that their negative frequency Fourier coefficients are equivalent to their positive frequency counterparts, any real-valued signal, such as $v(x, y, t)$, can be alternatively represented in its *analytical signal* form which consists of eliminating its negative frequency components. This alternative representation makes the extraction of the instantaneous amplitude and phase straightforward.

The analytical is commonly computed by using the Hilbert transform:

$$v_a(x, y, t) = v(x, y, t) + i\hat{v}(x, y, t) = a(x, y, t)e^{i\bar{\phi}(x,y,t)}, \tag{2}$$

where $\hat{v}(x, y, t)$ is the Hilbert transform of $v(x, y, t)$, $a(x, y, t)$ is the instantaneous amplitude, and $\bar{\phi}(x, y, t)$ the instantaneous phase wrapped in $[-\pi, \pi)$.

The use of the analytical signal assumes that the signal has a single frequency component at any time point. As can be seen in Fig. 1, this is not the case for the LFP recordings which have high power in the $10-45Hz$ band. In this work, we assume like in previous studies [7], that such "single-frequency" representation still captures important dynamics that are worth studying. Given that the power has a strong peak in $15 - 25$ Hz band, we filter the signal around 18 Hz with a bandwidth of $\pm$ 3 Hz, which strikes a reasonable balance between the accuracy of estimation and the ability to capture the richness of the wave phenomenon. Indeed, we observed in our initial exploration (not shown here) that with a narrower frequency band (e.g, $\pm$ 1Hz), the timeseries of estimated wave properties, in particular the direction of propagation timeseries, are less dynamic across time and give a less interesting perspective over the wave activity. Fig 3 illustrates the amplitude-phase decomposition of the signal of one trial. It is noticeable that the phase timeseries (i.e. bottom panel) alternates between segments of synchronized activity across channels and segments of less synchronization. This is related to the instantaneous amplitude, which indicates high synchrony and typically constructive addition in the oscillations across electrodes when it is high (e.g. at 600ms) and less synchrony when it is low (e.g. at -100ms relative to instruction cue).

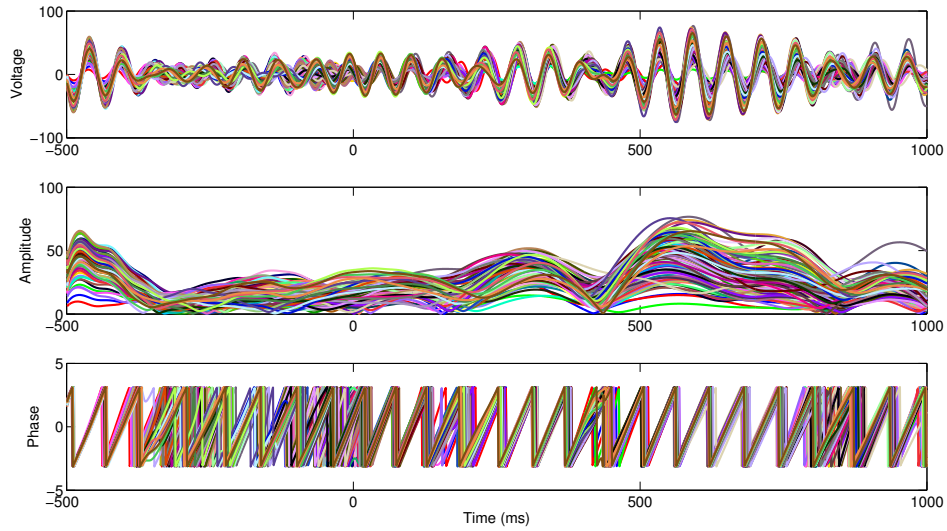

Figure 3: Analytical signals of all channels of one trial. **Top panel**: shows the voltage signal $v(x, y, t)$ filtered around 18 $\pm$ 3 Hz. **Middle panel** shows the associated amplitude $a(x, y, t)$ and the **Bottom panel** shows the phase $\bar{\phi}(x, y, t)$.

3.1.2 Phase Unwrapping

The wrapped phase $\bar{\phi}$ that we extract from the analytical signal varies between $[-\pi, \pi)$. Let us define the phase ϕ such that:

$$\begin{aligned} \bar{\phi} &= \mathcal{W}(\phi + \epsilon) \\ &= mod(\phi + \epsilon, 2\pi) - \pi \end{aligned} \tag{3}$$

where $\mathcal{W}$ is the wrapping operator and ϵ is random zero mean noise. Before being used as input to the linear regression method or the PGD method (where the 2π jumps would be misleading), the

phase needs to be tranformed back to ϕ, which can be interpreted as a spatio-temporal delay. In other words it needs to be *unwrapped*. Under the assumption that the sampling rate satisfies the Nyquist criterion in all dimensions, the absolute phase difference between any 2 adjacent points (spatially or temporally) should be no greater than π. Based on this principle, basic unwrapping procedures consist of navingating through the data points in any spanning tree structure and whenever the difference between adjacent data points is observed to be larger than 2π, add or subtract 2π to the second point [18]. When the timeseries is noisy, like in our phase timeseries, unwrapping becomes a hard problem because the unwrapping algorithm might erroneously add or subtract 2π or fail to do so, and the errors will propagate and corrupt the rest of the computation.

In this work, we implemented a quality guided unwrapping algorithm [19] which ensures that the noisy regions are unwrapped last and consequently the errors will remain local. It first ranks the points based on quality using a heuristic and then it unwraps the best parts of the data first and leaves the noisy parts to the end in order to minimize the propagation of the error. This is an appealing approach for our data because the noise is localized. Choosing the phase correlation between adjacent points has been previously shown to work on this type of data [17]. Fig. 4, shows the unwrapped phase of an example trial. Given that the signal is oscillating with an approximately constant period, the phase at every electrode is changing at a constant rate which is manifest in the slope of the unwrapped phase. As can be seen, the quality guided unwrapping helps containing the noise such that it remains localized.

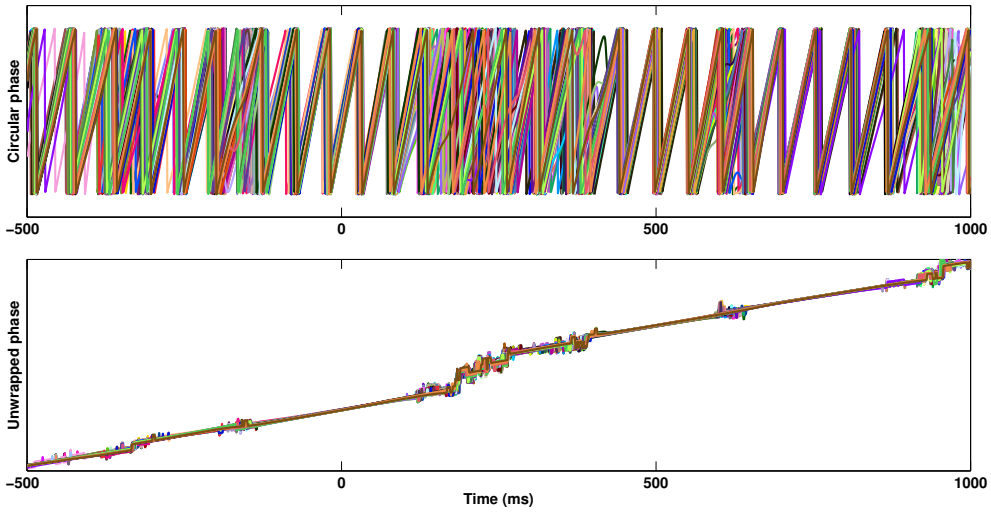

Figure 4: The phase of all channels of one trial. Note: the y-axis coordinates are not displayed. **Top panel**: shows the wrapped phase $\in [-\pi, \pi)$, **Bottom panel**: shows the unwrapped phase using the 3D quality guided unwrapping algorithm.

3.2 Linear wave analysis

3.2.1 Linear Regression Method

In the phase domain, the linear wave equation is equivalent to a plane model:

$$\phi(x, y, t) = k_x x + k_y y + \omega t + \phi_0 \tag{4}$$

which can be fit using linear regression [17]. As in [16], our results are based on regressing ϕ on x and y as independent variables and we estimate $\omega(t)$ separately by averaging the temporal derivatives across electrodes. In order to estimate the time-varying wave properties, we fit a series of linear regression models in overlapping time windows. For example, to estimate the wave properties at time t, we fit a model using data from $t - t_{win}$ to $t + t_{win}$, such that the total window size is

$2 \times t_{win} + 1$. The length of the time window is a parameter that we choose. In our initial exploration we looked at 5ms, 15ms and 31ms windows which all gave similar results. Below we use windows of 5ms.

We look at two diagnostic quantities associated with each fitted model: the p-value, which allows us to assess the statistical significance of the full model against the null hypothesis that $k_x = k_y = 0$; and R^2, which allows us to assess its goodness of fit. In Fig. 5 we see a typical times series of p-values of the fitted models of an example trial. Note that the $\log_{10}(p\text{-value})$ at all time points are extremely small suggesting that the dependence on x and y is statistically significant at all time points! This is due to the nature of the unwrapped phase in this problem which can be fit by a statistically significant plane at most time windows. However, more criteria are needed to distinguish a wave pattern from random linear pattern. The p-value alone is consequently clearly not adequate to distinguish actual waves from non-waves in this dataset.

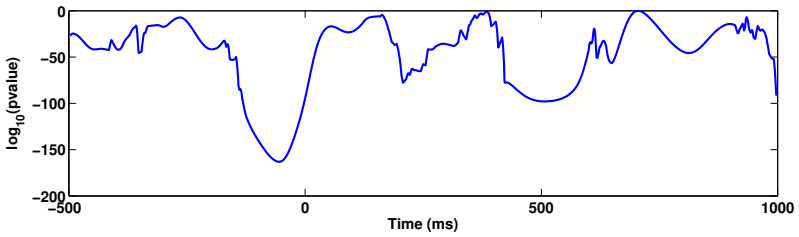

Figure 5: Full model p-value timeseries of one trial estimated using a regression window of 5ms.

What we are looking for is not merely that the linear approximation of the phase data be statistically significant. For a time segment to contain a wave, the linear pattern must also be strong across the electrodes meaning that the variability in the phase should be to a high degree explained by the variation in x and y. The R^2 is a measure of such strength. Formally, let

$$R^2 = \frac{E_v}{T_v}, \tag{5}$$

where $E_v = \sum_{x,y}(\hat{\phi}(x,y) - \bar{\hat{\phi}})^2$ is the variance explained by the linear fit, $T_v = \sum_{x,y}(\phi(x,y) - \bar{\hat{\phi}})^2$ is the total variance of the data, and $\bar{\hat{\phi}}$ is the mean phase over the observations. R^2 lies in $[0,1]$ where 0 means that none of the variance is explained by the fit, and 1 means that all of it is explained by the fit.

In this work, a "sufficiently" high R^2 is one of our criteria for defining a wave. In 5.1 we explain the two additional criteria corresponding to a "sufficiently" long duration and a "sufficiently" stable direction of propagation.

3.2.2 Phase Gradient Method

For the sake of comparison, we use the method developped to analyze this dataset in [2]. This method is based on numerically computing the instantaneous spatial and temporal derivatives at all electrodes. The average spatial phase gradient is then used to estimate the wavenumber vector. As seen in Fig. 2, the phase gradients of a perfect linear wave all point in the same direction. So a reasonable measure of the "linearity" of a wave, is found in the Phase Gradient Directionality (PGD) measure which computes an index of phase gradient alignment across the electrodes:

$$PGD(t) = ||\overline{\nabla\phi}||/\overline{||\nabla\phi||}, \tag{6}$$

with $PGD(t) \in [0,1]$, where 1 indicates perfect alignment and 0 indicates no alignment at all. [17] shows that this measure is similar results to the standard circular variance measure [20]. One drawback of this method is that it is based on instantaneous spatial derivative computations which exacerbate the noise.

Fig. 6 shows the estimation results from both R^2 (blue lines) and PGD (red lines). In the top panel an example R^2 timeseries can be seen to fluctuate between 0.1 and 0.8 at different times. The PGD estimates follow similar overall patterns but they are clearly noisier than R^2. This remains the

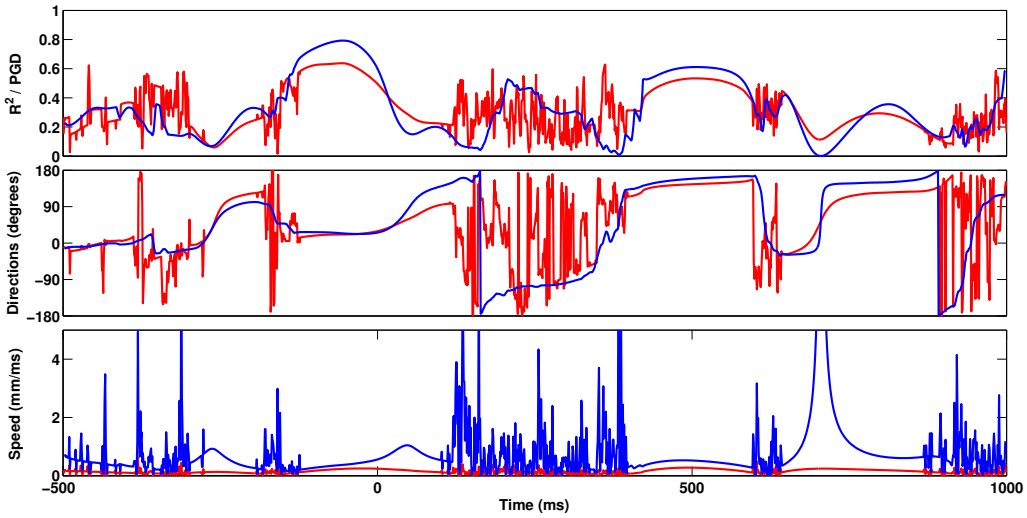

Figure 6: Wave estimates using regression (blue) and numerical derivatives (red) methods . **Top panel**: shows the R^2 and PGD timeseries, **Middle panel**: shows the direction of propagation timeseries in degrees, **Bottom panel**: shows the speed timeseries in m/sec.

case even if we smooth the PGD timeseries with a moving average filter of the same width as the regression window (not shown here). The middle panel shows the estimated directions of propagation. We can see that the 2 methods are relatively in agreement in less noisy segments, but again the PGD method suffers more from the noise and results in more high frequency fluctuations. In our initial exploration, we saw that direction estimates are quite stable as we vary parameters of the processing pipeline (including unwrapping algorithms, and regression windows sizes). Finally the bottom row shows the speed estimates. We can see that the regression estimates are systematically higher than the PGD estimates, and it is unsurprising because the wavenumber estimated in regression is smoother. However, a surprising difference in the two timeseries is the presence of the big, low frequency, peaks in the regression estimates but not in the PGD estimates. By looking at the regression residuals we found that those are usually associated with some systematic curvature in the phase data. For example, the phases at different times around 700 ms after the instruction cue all have the shape of a parabola as a function of the x and y coordinates which leads to a plane fit of very small slope k (and hence very high speed estimate computed as $\frac{\omega}{k}$). Note that the p-value at this point is also high in Fig. 5, indicating less evidence for k different than 0. This behavior might be a manifestation of the wave being reflected by some boundary and starting to propagate in the opposite direction. Because the PGD only looks at instantaneous average estimates of the wavenumber, this behavior is not captured. In both cases however, this window of time witnesses a change in the direction of propagation of around 180 degrees. Although this difference is interesting to investigate more thoroughly, note that these segments of time are associated with low R^2 and PGD values, and are therefore not of interest to us for the rest of the current analysis.

Now that we have timeseries of linear phase estimates, the question remains about which segments contain waves. We report this discussion to section 5 where the wave definition includes high R^2 or PGD values and stable direction of propagation estimates. In the rest of the document we will refer to R^2 or PGD as the test statistic when making the distinction is not relevant.

4 Array partitioning

In the rest of the work, we will regard the microelectrode array as a set of non-overlapping square patches. Specifically we have looked at patches of 3×3, of 4×4, and of 5×5 electrodes. The smaller the patch size the more localized are the dynamics we identify, however, as we will see in 6, the lower is the signal-to-noise ratio and the harder it is to distinguish real "wave segments" from noise. The different partitionings of the array are shown in Fig. 7. Every small square in the grid

represents an electrode, and the light brown parts delimit the patches. The blue parts are discarded from the analysis. Not that the 4×4 patch size results in the most separated patches among the three configurations.

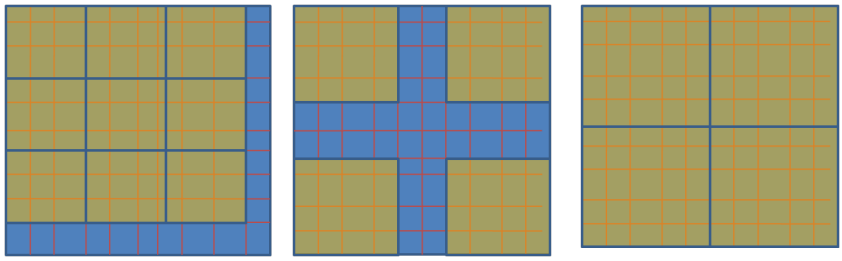

Figure 7: The different patch configurations overlayed on the 10×10 electrodes array. **Left panel**: 9 patches of size 3×3. They start at the top left electrode and partially cover the array. **Middle panel**: 4 patches of size 4×4, cover the four corners of the array. **Right panel**: 4 patches of size 5×5 cover the entire array.

5 Statistical Methods

Using the two methods above, we obtain a time series of test statistics and timeseries of direction and speed of propagation for every trial, patch, and experimental condition. However, the methods do not allow us to know which segments of the time series contain the statistically significant wave patterns that we are looking for. In the absence of a priori knowledge of the possible locations of these patterns, we have to conduct a hypothesis test at every time point, in every patch, trial and experimental condition and assess whether the observed statistic in that sample lets us or not to reject the null hypothesis (defined below) at some critical value α. This amounts to millions of tests and the need to control the Type I error rate through a multiple comparison procedure becomes essential. A Type I error is wrongly rejecting of the null hypothesis when it is actually true. Briefly, the challenge induced by such a large number of comparisons is that even if at individual samples the probability of wrongly rejecting the null is very small (0.05 for example), over the set of all comparisons we then expect to wrongly reject the null a large number of times (at tens of thousands of samples) only because of chance. Multiple comparison procedures are a means to control the Type I error rate across the set of comparisons instead of at each comparison separately. A corrolated challenge is then to control the Type II error rate, which consists of wrongly accepting the null. False Discovery Rate (FDR) procedures with improved sensivity have been developped to address this issue.

Below, we first define what we mean by a wave pattern in this work. Then, we describe the null hypothesis against which we compare our working hypothesis. Third we describe a permutation procedure called, depending on the source, the "suprathreshold cluster" test [21] or the "excursion" test [22] that lets us control for Type I error globally in every trial instead of at each time point. Finally, we conclude the section with the introduction of the FDR and the presentation of the Benjamini-Hoshberg procedure to control the type II error [23].

5.1 Wave segments definition

Our working hypothesis is that at least some trials contain segments of time that represent "wave patterns". We define a wave pattern to be a time segment of "sufficiently long" duration in which the phases are linearly organized across the electrodes of interest, indicated by a "sufficiently high" test statistic, and where the direction of propagation is "sufficiently stable". In this work, we consider a segment of at least 5ms duration to be sufficiently long and a cumulative sum of the absolute first difference in the direction timeseries of up to 15° to be a sufficiently small variation in direction of propagation. These values were chosen based on exploration and a discussion with a domain expert. The minimum time duration requirement is not essential, and is usually attained automatically when enforcing the stability of direction requirement. By looking at the distribution of direction of propagation at every time point of phase segments with strong linear patterns, we found that using a 15° threshold to delimit wave segments resulted in representative mean directions. We chose the test statistic threshold, called subsequently the primary threshold, above which the linear pattern is

sufficiently strong, to be the 99th percentile of the distribution of the test statistic under the null hypothesis. This means that using this threshold we retain the time points that have a 1 $\%$ probability of occuring if the null hypothesis was true. Once we have these three parameters, we can identify the segments of interest in our data (i.e. the segments that might contain waves), and we call them "the observed clusters" in the rest of the document . We compute their global p-values based on the excursion test and we select among them the ones that pass the FDR test procedure (described below) and consider them to be wave segments.

5.2 The null hypothesis

We define the null hypothesis to be that the spatial organization of the electrodes that generated the time series is completely random across all electrodes of the array. We simulate a draw from the null distribution by randomly shuffling the electrodes and computing the times series of the statistics of interest. Note that in this way we disrupt only the spatial organization across electrodes while preserving their temporal pattern and the overall cross-correlation structure (which is not affected by the order of the electrodes).

We compute the distribution of the test statistic under the null hypothesis for every experimental condition separately using the following randomization procedure:

1. Randomly pick a trial out of the set of trials associated with this condition.
2. Shuffle the electrodes across the entire array and then divide them into patches.
3. Process the phase, and estimate the wave parameters separately for every patch.

Repeat the above steps N_{1perm} times and in the end, the empirical distribution of the test statistic under the null, D_{null}, is represented by all the time points of all permutations and patches.

5.3 Excursion test procedure: Compute global p-values

The excursion test procedure allows us to control for the Type I error globally for an entire trial, instead of a sample by sample manner. Its rationale is the following: the permutation procedure described above computes a draw from the null distribution in an entire trial and so instead of looking at the distribution of test statistic at every time point in that trial, we could look at an appropriate maximal statistic that summarizes the entire trial. In that way, we reduce the multiple comparisons to a single comparison per trial and we can therefore control for the Type I error rate per trial. This is a weaker type of control than that provided by a sample-wise control, but it is the more appropriate for our purposes of finding activity that spreads smoothly spatio-temporally.

The maximal statistic used in an excursion test is an aggregate "cluster-level" statistic, where a cluster is defined to be a segment of time which has successfuly passed the three criterias of our wave definition (i.e. minimum R^2, minimum duration, and maximum direction variation) In this case we choose the aggregate cluster-level statistic to be the mean test statistic for that cluster, and the maximal statistic is the maximum over all clusters.

We compute the permutation distribution of the maximal aggregate statistic for every experimental condition separately using the following procedure:

1. Perform the randomization procedure described above and obtain N_{patch} statistic time-series and their associated direction timeseries.
2. Select all the time points where the statistic is higher than the primary threshold.
3. Cluster the selected points based on time adjacency.
4. Segment the clusters such that the cumulative absolute first difference in the direction time-series is less than a threshold D_{thresh} degrees.
5. Discard clusters with a duration less than 5 ms.
6. For every cluster compute its mean statistic.
7. Keep the maximum aggregate statistic over all the clusters per patch. If no clusters were found, set the maximum aggregate statistic to zero.

Repeat the above steps N_{2perm} times. In the end, the distribution of the maximum aggregate statistic under the null will be represented by the $N_{patch} \times N_{2perm}$ samples generated during this procedure.

As described earlier, we chose the primary threshold to be the 99th percentile of D_{null}. The primary threshold controls the power of the excursion test. The lower the threshold, the more clusters are retained in the first phase and the bigger (or longer) the found clusters are. In this case localized clusters are diluted by less intense adjacent activity. In contrast, if the primary threshold is high, the tendency favors smaller but more intense clusters and diffuse clusters are potentially missed. In our analysis we found that the requirement of directional stability played a bigger role in the duration of the found segments than increasing the primary threshold (e.g. to the 99.9 percentile).

Finally, the excursion test defines the global p-value of an observed cluster to be the probability that the maximal aggregate statistic under the null has a value more extreme than its observed aggregate statistic.

5.4 False Discovery Rate: Identify wave segments

After performing all the steps above, we are left with a total of $N_{segments}$ of the order of several thousands across all trials, patches, and experimental conditions, with their associated global p-values. This is still a large number of tests and the need to control for both Type I and Type II errors remains. Following [23], we choose to control the FDR criterion as opposed to controlling the more stringent Family Wise Error Rate (FWER) criterion. While the latter criterion controls the probability of making any Type I error and favors having a strong (sample-wise) control of the false error rate at the expense of a tremendous decrease in sensitivity of the test the former controls the expected number of errors favoring the sensitivity of the test at the price of a weaker control of the false error rate (or a more global control). In some problems, such as ours, controlling the FDR is more reasonable than controlling the FWER because a single mistake does not significantly affect our conclusions. In this work, we are interested in comparing distributions against each other, and so we are interested in controlling the proportion of mistakes, q, that we make rather than the probability of making any mistake.

For an expected proportion of mistakes, q, the procedure proposed by Benjamini-Hochberg [23] consists of the following: Let $p_1 \leq p_2 \leq \cdots \leq p_k \leq \cdots \leq p_{N_{segments}}$ be the ordered p-values of the $N_{segments}$. We find the largest k such that:

$$p_k \leq \frac{k}{N_{segments}} q, \tag{7}$$

and we reject the null hypothesis for all segments up to the kth one.

6 Results

Below we show the results of applying our methods on the dataset from [2]. Remember that at the instruction cue, the monkey is instructed to reach to one of 8 equally spaced directions. We label the different experimental conditions as {MVT1: 0°, MVT2: 45°, MVT3: 90°, MVT4: 135°, MVT5: 180°, MVT6: 225°, MVT7: 270°, MVT8: 315°}, where 0°points in the positive x-axis and the increase in angle is counter clockwise.

6.1 Null distribution and the choice of patch size

We performed $N_{1perm} = 1000$ permutations to compute the permutation distribution of the test statistics for each of the different patch sizes cases.

Table 1 lists the primary threholds computed for all experimental conditions for both R^2 and PGD, defined to be the 99th percentile of the permutation distribution of the corresponding test statistic. What to note first is that the thresholds across experimental conditions (columns) are always very similar. This is unsurprising, since we are expecting all of them to have the same kind of activity in terms of strength, but possibly with differences in their wave properties, in particular the directions of propagation. Second, we note that as the patch size increases, the threshold decreases (compare rows 1 and 6). This reflects the fact that for smaller patch sizes it is easier to fit a linear approximation

Row	PS	STAT	MVT 1	MVT 2	MVT 3	MVT 4	MVT 5	MVT 6	MVT 7	MVT 8
1	3	R^2	0.778	0.781	0.780	0.782	0.782	0.780	0.780	0.780
2	3	PGD	0.891	0.894	0.890	0.892	0.890	0.890	0.890	0.892
3	4	R^2	0.525	0.524	0.527	0.526	0.524	0.524	0.526	0.525
4	4	PGD	0.783	0.784	0.780	0.782	0.777	0.779	0.778	0.784
5	5	R^2	0.347	0.346	0.346	0.345	0.347	0.347	0.350	0.344
6	5	PGD	0.654	0.656	0.649	0.653	0.644	0.648	0.644	0.657

Table 1: Primary thresholds used to select time points where the test statistic is sufficiently strong. Shows the 99th percentile of the null distribution for both the R^2 and PGD statistics, across all experimental conditions and for the 3 different choices of patch configurations.

to the phase data than for larger ones. Third, we note that the R^2 is always smaller than the PGD threshold (compare rows 1 and 2), and the difference gap increases with patch size (compare rows 5 and 6). This is again unsurprising because in the regression method we are fitting the model using 5 consecutive time points while PGD estimates are computed only at individual time points. We therefore expect the R^2 statistic to be more sensitive to permutations than the PGD statistic.

Fig. 8 shows the distributions of the test statistics under the null with the associated distributions of the observed R^2 and PGD timeseries statistics for patch sizes of 3×3 (left subplot), 4×4 (middle subplot), and 5×5 (right subplot) in one example experimental condition (all others are similar). As the patch size increases, both the R^2 and PGD distributions become more peaky and have more mass to the left of the axis, however the R^2 distribution has a much sharper drop especially in the 5×5 patch size where the density beyond 0.5 is almost 0. In contrast, the PGD has probability mass beyond that point. As the patch size decreases, the distribution of the observed PGD shifts significantly towards one, but it is a less evident pattern in the case of the observed R^2.

The less overlap there is between the null and the observed distribution the easier it will be to detect the statistically significant patterns. At the same time, we are interested in analyzing local patterns across the array. We therefore choose to continue the analysis with patches of size 4×4 which have a sharper null distribution than patches of size 3×3 and are more localized than patches of size 5×5.

6.2 Global p-values and the choice of analysis method

We performed $N_{2perm} = 125000$ permutations to compute the distribution of the maximal aggregate statistic under the null hypothesis and $q = 0.05$ is the desired expected proportion of false rejections of the null. Fig. 9 shows all the observed clusters (across trials, patches, and experimental conditions) ordered by their p-values as estimated using the regression method (left panel) and using PGD (right panel). The first thing to note is that using R^2 more observed clusters are found that in the case of PGD, and more are retained by the FDR procedure (clusters below the green line are wave segments). This is unsurprising given the bigger overlap between the null and the observed distribution that we previously noted in the PGD case. We then choose to continue the analysis using the regression method estimates which proved to be more powerful in detecting linear patterns.

6.3 Different Activity across patches

Tables 2 and 3 summarize the statistically significant activity detected in patches averaged across experimental conditions. Table 2 shows the number of trials where each patch contained at least 1 segment of statistically significant activity (i.e. it is active). The top left patch (compare 1st column to last column) is always active, and in general the patches are active during most of the trials. Table 3 shows the average number of segments detected per patch per trial. Again, the top left patch is the one with the largest number of segments per trial, followed by the bottom left, bottom right, and then the top right patches. In the second row the associated average duration of a segment is similar across patches, with the bottom right patch having the largest average duration. These results are a first indication that the activity across the array is heterogeneous.

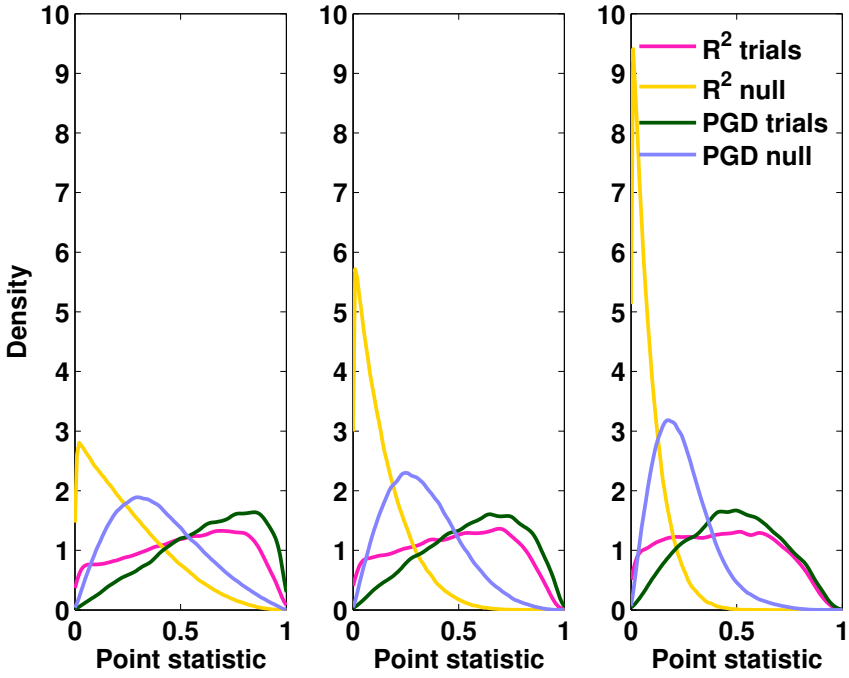

Figure 8: Distributions of the test statistics. **Left panel**: shows the null and observed distributions in 3×3 patches for both R^2 and PGD, **Middle panel**: shows the same for 4×4 patches, **Right panel**: shows the same for 5×5 patches.

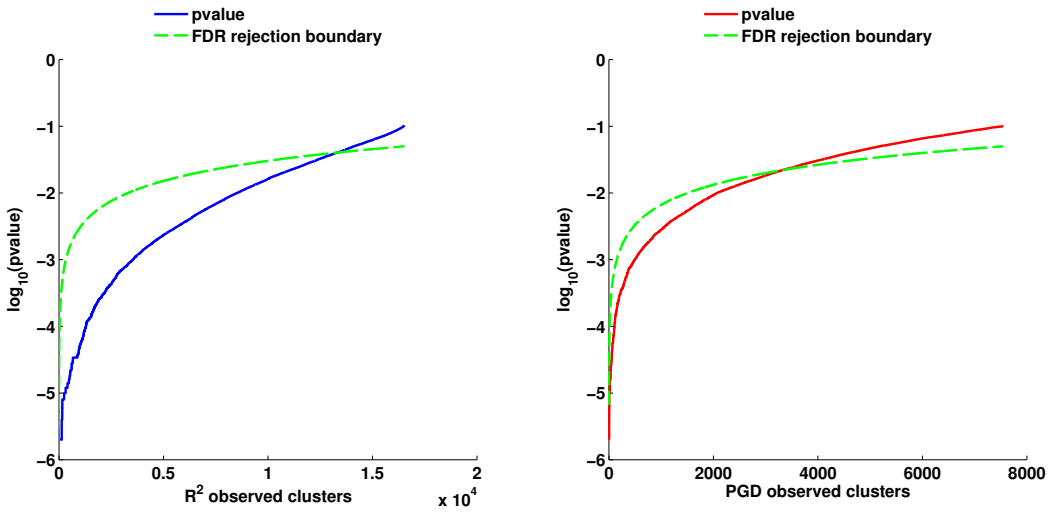

Figure 9: Observed cluster p-values for patches of size 4×4 and the FDR rejection boundary. The null is rejected for clusters below the green line. **Left panel**: shows the results of the regression method. **Right panel**: shows the results of the numerical derivatives method.

	Top left (1)	Bottom left (2)	Top right (3)	Bottom right (4)	Trial Numbers
MVT1	47	46	45	46	47
MVT2	46	45	44	46	46
MVT3	52	47	51	50	52
MVT4	41	40	40	41	41
MVT5	53	51	47	52	53
MVT6	61	61	58	61	61
MVT7	43	42	43	41	43
MVT8	48	47	46	48	48

Table 2: Number of active trials per patch for all experimental conditions. The first 4 columns show the numbe of trials where at least one wave segment is detected in each of the patches. The last column shows the total number of trials for each experimental condition.

	Top left (1)	Bottom left (2)	Top right (3)	Bottom right (4)
Avg. number of segments	12 ± 0.8	9 ± 1.1	6 ± 0.3	8 ± 1.2
Avg. segment duration	21.8 ms	21.5 ms	22.7 ms	24.1 ms

Table 3: Mean patch activity across experimental conditions. The first row shows the average number of selected patterns per patch per trial with its standard deviation, The second row shows the average total duration of selected patterns per patch per trial.

6.3.1 Mean Direction across experimental conditions

In Fig. 10 we show the distribution of the directions of the significant patterns that start after the instruction cue. We overlay the mean direction in red, with a radius equal to the resultant vector length. We lay the results from all experimental conditions in a square, where every block, represents an experimental condition, and is placed corresponding to the target direction. The subplots in every block represent the patches of size 4×4 (4 patches) and we number the patches in row major format (e.g., patch 2 corresponds to the patch in the second row and 1st column).

We first note that there is considerable heterogeneity across the patches. Second, we note that the overall pattern per patch are relatively stable across experimental conditions in comparison to the differences across patches. For example, the directions of the top left patches are usually bimodal and with a stonger bias toward the upper left quadrant, while the directions of the bottom right patches are consistently found in the bottom half of the directions axes. The numbers on top of each rose plot corresponds to the number of segments whose mean directions are shown. We also investigated the differences between the directions in a same patch location across experimental conditions. While we have found some statistically significant differences (with p-values < 0.001), we found that the effect size was too small to be interesting to elaborate on this point.

7 Discussion

In this work, we have presented a statistical methodology for the analysis of linear wave patterns in LFP recorded by a 10×10 Utah grid array. The main steps of this procedure were first the definition of a null hypothesis with a way to simulate draws from it, second an excursion test procedure which controls globally for the Type I error in a trial, and third an FDR procedure which increases the sensitivity of the test while controlling the proportion of false rejections of the null. The particular null hypothesis that we defined in this work was that the phase patterns were random across the entire array. Using this procedure we analyzed the local patterns of statistically significant linear waves by dividing the array into smaller 4×4 patches and focusing in particular on the mean directions of propagation of wave segments. Our results suggest that there is spatial heterogeneity across patches both in terms of number of patterns found across different patches and the distributions of their mean directions of propagations. However, we were not able to find systematicity in how the patterns change as a function of location inside the array or as a function of target direction.

Once the null hypothesis defined, the identified wave patterns depended on a few choices we made, the main one being how variable we allowed the direction of propagation of a single wave segment to be. However, we found that the identified relative numerosity of patterns and directions of propgations were robust to a variety of other parameter choices which suggest that the patterns we found

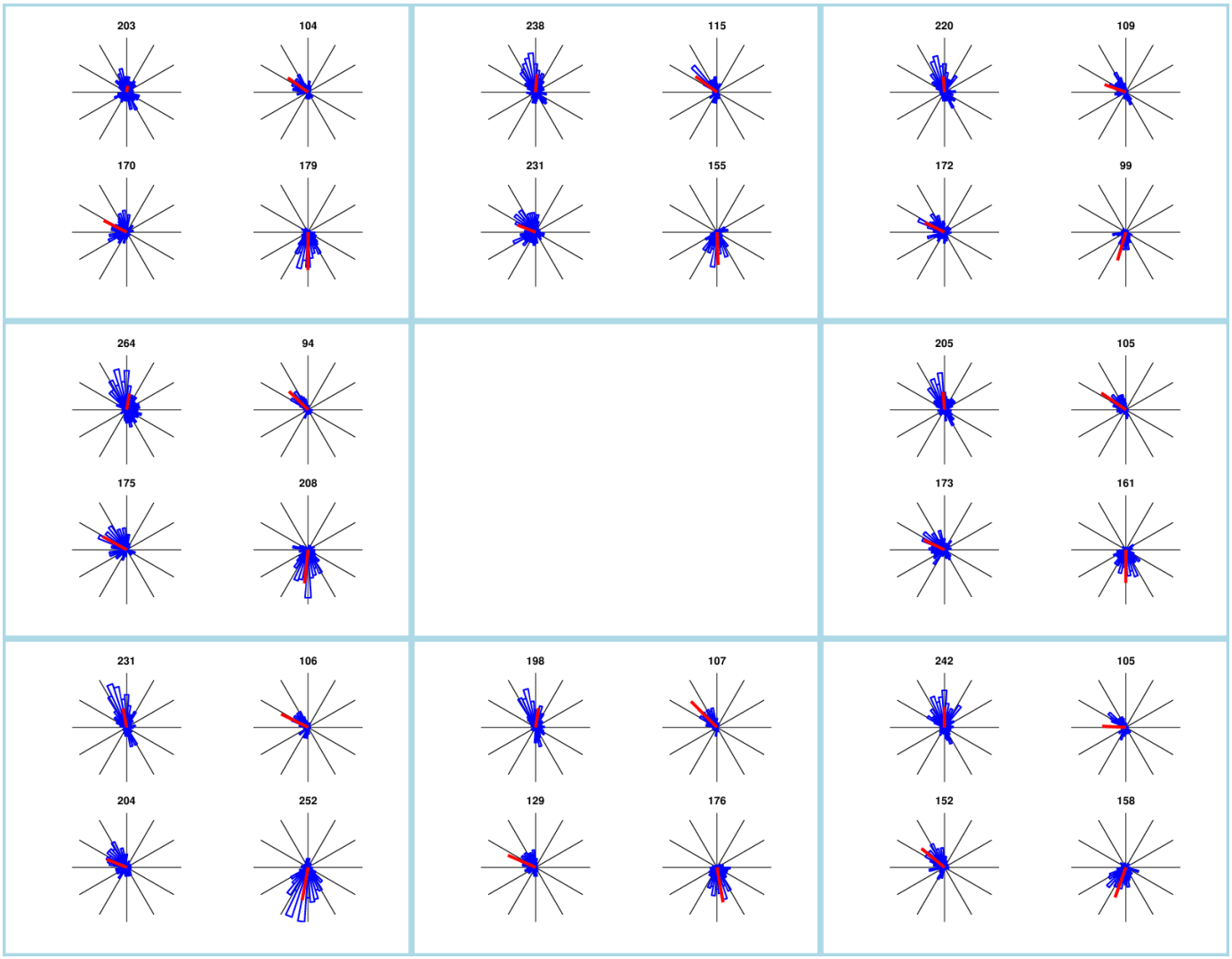

Figure 10: Direction rose plots of significant patterns in all patches and all experimental conditions in 4×4 patches, taken within 500ms after the instruction cue. The experimental conditions are organized based on the direction of the target movement, and the patches are orgaznized based on their order on the array. We indicate the number of wave segments for which the mean angle is included in the rose plot on top of every patch.

are real and quite strong. Our statistical analysis demonstrated important differences between the regression and the numerical differentiation methods. In particular, due to the noisier estimates by the PGD method, it detected a smaller number of statistically significant patterns. In both cases, we found that the estimation of speed of propagation was the most unreliable and diverged significantly between the two methods.

Future improvements to this line of work concern improving the phase unwrapping by combining it with a 3-dimensional smoothing algorithm. Our initial exploration (not presented here) pointed us in a couple of directions to do that. Either first use a 3-dimensional nonparametric smoothing algorithm in the phase domain and then unwrap the smoothed phase, or instead use a Kalman smoothing algorithm that performs both operations simultaneously. We expect that this additional step will improve the speed estimation which has proved to be the most challenging. In addition, a more specific definition of the null (beyond complete randomness) might improve robustness of the results.

The main limitation of this approach is the assumption that the phenomenon is contained in a narrow-band frequency signal. As we have noted before, given that the signal is broad band it is likely that a better model to represent it is a superposition of several linear waves instead of only one. A future direction of work could be that of modeling the entire wide-band signal, using periodic autoregressive models for example. Another assumption we make is that the real difference between any adjacent phase points (both in time and space) is less than π. This assumption is commonly made to do the unwrapping, however more investigation is needed to know if it is valid in our setting.

This work is a first step towards investigating the morphology of linear waves in the motor cortex. A direct extension of the analysis will be to look at the different wave properties (e.g. speed, timing, duration, etc.) in relation to behavior and exact location of the electrodes in the cortex as well as to analyze more datasets.

References

[1] Paul L Nunez and Ramesh Srinivasan. A theoretical basis for standing and traveling brain waves measured with human eeg with implications for an integrated consciousness. *Clinical Neurophysiology*, 117(11):2424–2435, 2006.

[2] Doug Rubino, Kay A Robbins, and Nicholas G Hatsopoulos. Propagating waves mediate information transfer in the motor cortex. *Nature neuroscience*, 9(12):1549–1557, 2006.

[3] Andrea Benucci, Robert A Frazor, and Matteo Carandini. Standing waves and traveling waves distinguish two circuits in visual cortex. *Neuron*, 55(1):103–117, 2007.

[4] Antonia Reimer, Peter Hubka, Andreas K Engel, and Andrej Kral. Fast propagating waves within the rodent auditory cortex. *Cerebral Cortex*, page bhq073, 2010.

[5] Evgueniy V Lubenov and Athanassios G Siapas. Hippocampal theta oscillations are travelling waves. *Nature*, 459(7246):534–539, 2009.

[6] Nicholas G Hatsopoulos, Leonel Olmedo, and Kazutaka Takahashi. Proximal-to-distal sequencing behavior and motor cortex. *Motor Control: Theories, Experiments, and Applications*, pages 159–176, 2010.

[7] Kazutaka Takahashi, Sanggyun Kim, Todd P Coleman, Kevin A Brown, Aaron J Suminski, Matthew D Best, and Nicholas G Hatsopoulos. Large-scale spatiotemporal spike patterning consistent with wave propagation in motor cortex. *Nature communications*, 6, 2015.

[8] Gregory J Jarmer, Eric B Flynn, and Michael D Todd. Dispersion curve estimation via phased array beamforming methods. *Journal of Intelligent Material Systems and Structures*, page 1045389X13494930, 2013.

[9] Melvin J Hinich. Frequency–wavenumber array processing. *The Journal of the Acoustical Society of America*, 69(3):732–737, 1981.

[10] Matthias Ohrnberger, Estelle Schissele, Cecile Cornou, Sylvette Bonnefoy-Claudet, Marc Wathelet, Alexandros Savvaidis, Frank Scherbaum, and Denis Jongmans. Frequency wavenumber and spatial autocorrelation methods for dispersion curve determination from ambient vibration recordings. In *Proc. of 13th World Conf. on Earthquake Engineering, Vancouver, BC, Canada*, pages 1–6, 2004.

[11] Marshall D Earle. Nondirectional and directional wave data analysis procedures. *NDBC Tech. Doc*, pages 96–01, 1996.

[12] Hamutal Slovin, Amos Arieli, Rina Hildesheim, and Amiram Grinvald. Long-term voltage-sensitive dye imaging reveals cortical dynamics in behaving monkeys. *Journal of neurophysiology*, 88(6):3421–3438, 2002.

[13] Xiaoying Huang, Weifeng Xu, Jianmin Liang, Kentaroh Takagaki, Xin Gao, and Jian-young Wu. Spiral wave dynamics in neocortex. *Neuron*, 68(5):978–990, 2010.

[14] G Bard Ermentrout and David Kleinfeld. Traveling electrical waves in cortex: insights from phase dynamics and speculation on a computational role. *Neuron*, 29(1):33–44, 2001.

[15] David M Alexander, Peter Jurica, Chris Trengove, Andrey R Nikolaev, Sergei Gepshtein, Mikhail Zvyagintsev, Klaus Mathiak, Andreas Schulze-Bonhage, Johanna Ruescher, Tonio Ball, et al. Traveling waves and trial averaging: The nature of single-trial and averaged brain responses in large-scale cortical signals. *Neuroimage*, 73:95–112, 2013.

[16] Evgueniy V Lubenov and Athanassios G Siapas. Hippocampal theta oscillations are travelling waves. *Nature*, 459(7246):534–539, 2009.

[17] Bronwyn Woods. Spatio-temporal patterns in multi-electrode array local field potential recordings. *arXiv preprint arXiv:1501.00230*, 2015.

[18] Kazuyoshi Itoh. Analysis of the phase unwrapping algorithm. *Appl. Opt*, 21(14):2470, 1982.

[19] R Cusack and N Papadakis. New robust 3-d phase unwrapping algorithms: application to magnetic field mapping and undistorting echoplanar images. *Neuroimage*, 16(3):754–764, 2002.

[20] Kanti V Mardia and Peter E Jupp. *Directional statistics*, volume 494. John Wiley & Sons, 2009.

[21] Thomas E Nichols and Andrew P Holmes. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human brain mapping*, 15(1):1–25, 2002.

[22] Yang Xu, Gustavo P Sudre, Wei Wang, Douglas J Weber, and Robert E Kass. Characterizing global statistical significance of spatiotemporal hot spots in magnetoencephalography/electroencephalography source space via excursion algorithms. *Statistics in medicine*, 30(23):2854–2866, 2011.

[23] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 289–300, 1995.