1 **EXTRACTING INFORMATION IN SPIKE TIME PATTERNS WITH WAVELETS AND**

2 **INFORMATION THEORY**

3 Vítor Lopes-dos-Santos[1,2], Stefano Panzeri[3,4], Christoph Kayser[5,6], Mathew E. Diamond[7], Rodrigo Quian

4 Quiroga[2*]

5 [1]Brain Institute, Federal University of Rio Grande do Norte, Natal, Brazil.

6 [2]Centre for Systems Neuroscience, University of Leicester, Leicester, UK.

7 [3] Center for Neuroscience and Cognitive Systems @UniTn, Istituto Italiano di Tecnologia, 38068
8 Rovereto, Italy.

9 [4]Max Planck Institute for Biological Cybernetics, 72076 Tübingen, Germany.

10 [5]Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, UK.

11 [6]Bernstein Center for Computational Neuroscience, Tübingen, Germany

12 [7] Tactile Perception and Learning Lab, International School for Advanced Studies, Trieste, Italy.
13
14 * Corresponding author:

15 Centre for Systems Neuroscience,

16 University of Leicester

17 Leicester, LE1 7QR,

18 United Kingdom

19 (p) +44 116 252 5200

20 (e) rqqg1@le.ac.uk

21

22 **KEY WORDS:** Neuronal coding, temporal coding, wavelets, information theory, decoding.

23 **RUNNING HEAD:** NEWMETHOD TO EXTRACT INFORMATION IN SPIKE TIME PATTERNS.

24

ABSTRACT

We present a new method to assess the information carried by temporal patterns in spike trains. The method first performs a wavelet decomposition of the spike trains, then uses Shannon information to select a subset of coefficients carrying information and, finally, assesses timing information in terms of decoding performance – the ability to identify the presented stimuli from spike train patterns. We show that the method allows: i) a robust assessment of the information carried by spike time patterns even when this is distributed across multiple time-scales and time-points, ii) an effective denoising of the raster plots that improves the estimate of stimulus tuning of spike trains, and iii) an assessment of the information carried by temporally coordinated spikes across neurons. Using simulated data we demonstrate that the Wavelet-Information (WI) method performs better and is more robust to spike time-jitter, background noise and sample size than well-established approaches, such as principal component analysis, direct estimates of information from digitized spike trains or a metric-based method. Furthermore, when applied to real spike trains from monkey auditory cortex and from rat barrel cortex, the WI method allows extracting larger amounts of spike timing information. Importantly, the fact that the WI method incorporates multiple time-scales makes it robust to the choice of partly arbitrary parameters such as temporal resolution, response window length, number of response features considered, or the number of available trials. These results highlight the potential of the proposed method for accurate and objective assessments of how spike timing encodes information.

INTRODUCTION

The importance of precise spike timing in carrying meaningful information has attracted much attention (Quiroga and Panzeri 2009; Rieke et al. 1999). Does the temporal structure of spike trains provide information beyond the total spike count, or does it merely reflect noise? According to the "rate coding" view, neurons represent stimuli solely by the rate of firing within an encoding time window (Adrian and Zotterman 1926; Shadlen and Newsome 1994). In contrast, according to the "temporal coding" view, the time structure of the responses conveys additional information not provided by the total spike count (de Ruyter van Steveninck et al. 1997; Optican and Richmond 1987; Richmond and Optican 1987; Victor and Purpura 1996).

Experimental evidence accumulated over the last three decades has suggested that precise spike patterns - on the scale of milliseconds - do indeed convey information not available in rate codes (Arabzadeh et al. 2006; de Ruyter van Steveninck et al. 1997; Di Lorenzo et al. 2009; Eckhorn and Popel 1975; Foffani et al. 2009; Fontanini and Katz 2006; Kayser et al. 2010; Laurent et al. 1996; Montemurro et al. 2007; Panzeri et al. 2010; Panzeri et al. 2001; Quiroga and Panzeri 2009; Richmond and Optican 1987; Victor 2000). For this, a straightforward way to assess the significance of spike timing has been to represent spike trains as sequences of '0's and '1's denoting the absence or the presence of a spike in post-stimulus time bins and then, using the formalism of information theory, evaluate whether the information about stimulus identity carried by such patterns is significantly larger than the information carried by spike counts alone (de Ruyter van Steveninck et al. 1997; Kayser et al. 2009; Panzeri et al. 2001; Strong et al. 1998). However, this approach leads to a combinatorial explosion (the "curse of dimensionality"), because the number of possible response patterns increases exponentially with the number of bins (Panzeri et al. 2007). Thus, for an experimentally feasible number of trials, this limits the precision of the temporal patterns to be studied (i.e. the size of the time bin) and the length of the response considered.

69   A solution to the combinatorial explosion problem is to reduce the dimensionality of the spike

70   trains. To this end, a well-known approach is to compress the neural responses into a small number of

71   features using Principal Component Analysis (PCA). By using this method, Richmond and Optican (1987)

72   showed that time patterns in responses from neurons in the macaque inferior temporal cortex could

73   disambiguate visual stimuli that could not be distinguished by firing rate alone (Optican and Richmond

74   1987; Richmond and Optican 1987). Despite the value of this application, the PCA-based time

75   decomposition has two main caveats. First, PCA represents directions of maximum variance, which are

76   not necessarily the directions with largest information. Second, PCA coefficients are not localized in time

77   and may not capture sources of information that are precisely localized at one or a few restricted post-

78   stimulus time points (Panzeri et al. 2001), or even encoded at multiple temporal scales (Fotowat et al.

79   2011; Harvey et al. 2013; Kayser et al. 2009; Panzeri et al. 2010). Here we propose another type of

80   dimensionality reduction that is able to capture time-localized information encoded at multiple time

81   scales (Figure 1). The method combines wavelet decomposition and information theory to first identify

82   features in the spike patterns carrying relevant information, and then use these features to quantify the

83   amount of sensory encoding carried by these responses using a decoding approach. We validate the

84   method on simulated spike trains and compare its performance to that obtained with PCA, direct

85   estimations of information from digitized neuronal responses and a widely used Metric-Space (MS)

86   method (Victor and Purpura 1996). Results on simulated data demonstrate that the Wavelet-

87   Information (WI) method is more robust and extracts more spike timing information than previous

88   methods for a wide range of background firing rates and inter-trial jitters. The advantages of the WI

89   method are confirmed by evaluating its performance with experimental data from the monkey auditory

90   cortex and the rat somatosensory cortex. Additionally, we show that the same approach can be used to:

91   i) denoise spike trains, providing a more robust quantification of the stimulus selectivity and ii) assess

92   and visualize the information carried by the synchronous firing of neurons.

93

MATERIALS AND METHODS

WAVELET DECOMPOSITION

The wavelet transform is the inner product of a signal with dilated and translated versions of a

wavelet function (Mallat 2008; Strang and Nguyen 1996).Formally, given a signal $x(t)$ and a wavelet

function $\psi_{a,b}(t)$ the continuous wavelet transform (CWT) is defined as:

$$W_\psi X(a,b) = \left\langle x, \psi_{a,b} \right\rangle$$, with

$$\psi_{a,b} = |a|^{\frac{-1}{2}} \psi(\frac{t-b}{a})$$,

where $a, b \in \mathcal{R}$ are the scale and translation parameters, respectively. The translation parameter

changes the location of the wavelet function, while the scaling parameter dilates or compresses it. The

correlation of the signal $x(t)$ with the dilated (contracted) versions of the wavelet $\psi_{a,b}(t)$ gives the low

(high) frequency components. The CWT is very redundant and, without any loss of information, it is

practical to define the wavelet transform only at discrete scales $a_j = 2^j$ and times $b_{j,k} = 2^j k$, which is

called the dyadic wavelet transform (DWT). The DWT is non-redundant, in the sense that from *N* data

points we obtain *N* wavelet coefficients, each of them representing the amount of activity of the original

signal at a specific time and scale. Further, patterns in the signal with different frequency and time

localizations are represented by specific wavelet coefficients. The DWT can be computed using a

hierarchical and very efficient algorithm called *multi-resolution decomposition* (Mallat 1999). This

algorithm successively divides the signal into coarse approximations and details at different scales. The

112 end result is the decomposition of the original signal into a series of detail scales and a final

113 approximation, corresponding to the time-localized activity in different frequency bands.

114 Starting from the binned spike trains, in this study we implemented a 5-scale dyadic wavelet

115 decomposition using Haar wavelets, which is a square function that is ideally suited to identify local

116 contrasts at different scales. The spike trains were always binned with 1-ms windows unless stated

117 otherwise.

118

119 *SELECTION OF WAVELET COEFFICIENTS*

120 From the total set of wavelet coefficients, equal to the number of bins in the spike trains, we

121 selected a subset of coefficients based on their mutual information with the stimuli, defined as

122 (Shannon 1948):

123 $$I_{S,w_c} = \sum_{S,w_c} P(S,w_c) \log_2 \frac{P(S,w_c)}{P(S)P(w_c)},$$

124 where S is the set of stimuli and w_c is the set of values of a wavelet coefficient c. The significance of

125 the information given by each coefficient I_{S,w_c} was established based on surrogate testing: for each

126 coefficient we calculated a distribution of information values obtained by shuffling trials (i.e.,

127 randomizing trial-stimulus relations) 20 times. Surrogate distributions were calculated separately for

128 each decomposition level (information values obtained from coefficients of the same levels were

129 combined) and the 95 percentiles of each distribution were used as statistical thresholds (horizontal

130 dashed lines in Figure 1C). To avoid having too many features with significant information, if more than

131 25 coefficients were significant, we used the 25 with largest unbiased information (the same restriction

132 was applied to the PCAinfo method). In this context, unbiased information was defined as the difference

133 of the direct measure of information and its corresponding statistical threshold computed from shuffling

134   the trials. Additionally, if none of the coefficients crossed the statistical threshold, we used the two with

135   the largest information.

136

137   *STIMULUS DECODING AND INFORMATION ESTIMATION FROM CONFUSION MATRICES*

138   In order to estimate information in a set of features, we used a cross-validated (leave-one-out)

139   naïve Bayesian decoder to assign the response on each trial in the testing set to a given stimulus, which

140   gives a lower bound of the information available in the spike trains (Quiroga and Panzeri 2009).

141   Decoding performance was computed as the proportion of correct predictions. The conditional density

142   functions and the selection of features were based solely on the training trials. For comparison, we also

143   used Linear Discriminant Analysis (Fisher 1936; Quiroga et al. 2007) and Nearest Neighbors classifiers (in

144   which case, features were assigned z-scores in order to avoid scaling problems) and obtained virtually

145   the same results.

146   In cases where the linear decoder introduced systematic errors (Figure 7), we computed the

147   mutual information between the actual and the predicted stimuli from the confusion matrices:

$$I(S, S^P) = \sum_{S, S^P} P(S, S^P) \log_2 \frac{P(S, S^P)}{P(S)P(S^P)}$$

148   ,

149   where S is the set of actual stimuli presented to the decoder and S^P is the set of predicted stimuli by

150   the decoder. In order to correct for the upward limited-sampling bias in the information estimate, we

151   used the quadratic extrapolation procedure described elsewhere (Panzeri et al. 2007; Strong et al. 1998)

152   and implemented in Magri et al. (2009).

153

154

7

155

METRIC SPACE METHOD

We also compared the WI method with the Metric Space (MS) approach (Victor and Purpura 1996), which clusters responses based on a distance metric between spike trains. This distance is defined as the minimum "cost" of converting a spike train into another one by deleting, inserting or moving spikes. The cost of deleting or inserting a spike is always set to 1, and the cost of moving a spike per unit of time is given by the free parameter q (expressed in units of 1/ms), which has to be optimized for each dataset. Thus, when q is zero, moving a spike is free and therefore only the spike count is taken into consideration. As q is increased, more weight is given to the precise timing of the spikes. Note that since moving a spike by $1/q$ ms has the same cost of deleting it, $1/q$ defines the temporal precision of the analysis. With the MS method, we classified trials using a nearest neighbor decoder - more specifically, we assigned each tested trial to the class of its nearest neighbor in the training set.We systematic varied q from 0.001/ms to 524/ms, in half octave intervals. Only representative q values are reported. For computing the spike train distances we used a MATLAB function available at www-users.med.cornell.edu/~jdvicto/spkdm.html.

170

SPIKE TRAIN DENOISING

To visualize spike patterns containing information, we adapted the WI method to denoise the spike trains by: (1) computing the wavelet decomposition of the mean PSTH of each stimulus, (2) denoising the mean PSTHs by reconstructing them using only the wavelet coefficients with significant information, (3) setting to zero the denoised PSTHs values below a threshold at 1 s.d. of the absolute values (taken from the denoised PSTHs of all stimuli), and (4) using the denoised and thresholded PSTHs as *masks*, preserving only spikes in bins passing this threshold. This effectively preserves spikes

178   conveying stimulus information and deletes the others. We remark that although the selection of

179   informative wavelet coefficients was the same for all stimuli, the *masks* were different for each stimulus.

180

181   *ESTIMATION OF INFORMATION IN CORRELATED SPIKE PATTERNS*

182   To estimate the information given by the correlated firing of pairs of neurons *(i,j)*, we computed

183   the wavelet decomposition for each trial *n* and calculated the normalized distance between the values

184   of a corresponding wavelet coefficient (i.e. considering a specific time location and scale) in both

185   neurons, as:

186    $$D_{c,n}^{i,j} = \frac{\left|w_{c,n}^i - w_{c,n}^j\right|}{\left|w_{c,n}^i\right| + \left|w_{c,n}^j\right|} \ ,$$

187   where $w_{c,n}^i$ is the value of wavelet coefficient *c* of neuron *i* at trial *n*. Analogous to the procedure

188   described in Figure 1C, we then selected the distances $D_c^{i,j}$ that had significant information about the

189   stimuli and used these distance values for decoding. In other words, we implemented the same

190   procedure as before but using the distances $D_c^{i,j}$ between the wavelet coefficients of each neuron

191   instead of the value of the coefficients themselves.

192

193   *EXPERIMENTAL DATA*

194   *Monkey A1 data*

195   As described in a previous work (Kayser et al. 2010), neural activity was recorded from caudal

196   auditory cortex (mainly areas A1 and caudal belt) of three alert animals using multiple microelectrodes.

197   The data was high-pass filtered (4 Hz), amplified (Alpha Omega system) and digitalized at 20.83 kHz.

198 Recordings were performed in a dark and anechoic booth while animals passively listened to acoustic

199 stimuli. The sound stimulus consisted in a 40-second sequence of pseudorandom tones ("random

200 chords"). This sequence was generated by presenting multiple tones (125-ms duration) in different

201 sequences (12 fixed frequency bins per octave), with each tone frequency appearing (independently of

202 the others) with an exponentially distributed inter-tone interval (range 30-1000 ms, median 250 ms). To

203 estimate spectrograms of the acoustic stimulus, the signal was convolved with complex Morlet wavelets

204 with central frequencies ranging from 20 to 1600 Hz. Then, the instantaneous amplitude of each

205 "frequency" was computed as the norm of the complex values. The z-scored instantaneous amplitudes

206 were used for computing spike triggered averages for reverse correlation.

207 *Rat S1 data:*

208 As described in previous works (Lebedev et al. 2000; Panzeri et al. 2001), recordings in the

209 somatosensory cortex of adult Wistar rats were performed with an array of six tungsten

210 microelectrodes. Neurons in barrel-columns C1-3, D1-3 and E1-3 were recorded, while their

211 corresponding whiskers were stimulated individually. The stimulus was an up-down step function of

212 80 μm amplitude and 100-ms duration, delivered 48 times for each vibrissa with a 1 second inter

213 stimulus interval. Neuronal activity was amplified and band-pass filtered in the range 300-7500 Hz. Spike

214 waveforms were digitized at 25kHz (Datawave Discovery, Boulder, CO).

215

216 **RESULTS**

217 *OUTLINE OF THE WAVELET INFORMATION FRAMEWORK*

218 The first step of the WI method is to convolve the spike train responses (to repeated

219 presentations of a set of stimuli; Figure 1A) with Haar wavelets (Figure 1B; see Methods). Thus, each

220   trial is decomposed into a set of wavelet coefficients representing local spike patterns at different time

221   scales. In order to identify wavelets carrying meaningful information, we then compute the mutual

222   information between each coefficient and the stimuli (Figure 1C). Information values are compared to

223   distributions constructed by stimulus label permutations (see Methods). Then, the wavelet coefficients

224   with significant information are used to represent the data. We used a decoding approach to quantify

225   the performance of the WI method (and other methods for comparison) in extracting stimulus

226   information. Figure 1D shows confusion matrices of naïve Bayesian decoders trained to classify

227   responses with time patterns as in Figure 1A, either with spike counts or with the selected wavelet

228   coefficients. As expected, by construction, the performance with the WI method clearly outperformed

229   the one obtained with spike counts.

230

231   *PERFORMANCE WITH SIMULATED DATA*

232   We used simulated data to quantify the performance of the WI method and compared it to

233   other approaches. The simulated data consisted of a set of 200-ms responses to four hypothetical

234   stimuli created with a two-step procedure: (1) a specific spike time pattern (a sequence of pre-defined

235   spike times) was assigned to each stimulus and inserted into the response of a given trial with a random

236   shift within a window centered at their original time; (2) background activity was generated

237   independently for each trial following a Poisson process with a given mean rate and was then added to

238   these patterns.

239   The example in Figure 2A illustrates the ability of the WI method to extract information at

240   different time scales. A relatively precise spike timing distinguishes the first two stimuli (at the bottom;

241   generated using a jitter of 0.5 ms) and a pattern at a coarser scale distinguishes the remaining two (at

242   the top; generated using a jitter of 8 ms). The mean background rate was 8 spikes per second. The right

243   panel of Figure 2B shows the outcome of the WI method using a fine (1 ms) binning of the data. The

244   mean decoding performance was close to perfect (0.975), thus indicating the ability of the method to

245   capture time patterns at different scales. To further understand this result, in Figure 2C we show the

246   decoding outcomes obtained when considering coefficients of each wavelet scale separately. While the

247   coefficients from the coarser scales (scale 4, 5 and last approximation) could distinguish the coarse time

248   patterns of stimuli 3 and 4, the high frequency coefficients (scales 1, 2 and 3) distinguished the more

249   precise patterns of stimuli 1 and 2. For scale 1 no coefficient crossed the statistical threshold and we

250   therefore used the two with largest information.

251   For comparison with another dimensionality reduction method, we applied the PCA based

252   developed by Richmond and Optican (1987). For this, we computed the principal components (PC) from

253   the spikes trains binned either with 1-ms or with 8-ms windows (vertical lines in Figure 2A). As with the

254   WI method, we estimated the time pattern information by decoding stimuli based on the scores of the

255   four PCs with largest variance. As shown in Figure 2B, neither the 1- or 8-ms bins could capture the

256   information at both time scales. In particular, with the 1 ms binning the decoder could distinguish

257   between stimuli 1 and 2 but not between 3 and 4, given that pattern of these 2 stimuli were scattered

258   across several bins. Likewise, when using the 8 ms bins the decoder could distinguish between stimuli 3

259   and 4, but this gridding was too coarse to distinguish between stimulus 1 and 2.

260   To test the WI method in scenarios mimicking different recording conditions, we generated 3

261   examples (Figure 3) including patterns with different time localizations, precision and complexity, and

262   we systematically varied the background firing rate (2 to 64 Hz) and the time jitter (2 to 64ms). Left

263   panels show simulations with low background rate and jitter (4Hz and 2ms, respectively), while right

264   panels show simulations of the same patterns but with larger baseline firing and jitter values (16Hz and

265   8ms).

266   We further implemented a similar PCA-based approach but selecting the principal components

267   with the highest information. Also, we calculated performance using the whole binned responses (i.e.,

268   with no reduction of dimensionality). We then compared these 'time pattern' strategies to a total spike

269   count decoding, simply summing the total number of spikes of each trial. The selection of features was

270   always performed on a set of trials used to train the decoder (training set), and then performance was

271   evaluated in a different set of trials (test set).

272   Figure 4A displays results for different jitters and background rates for Example 1 of Figure 3

273   (using 15 trials per stimulus for training and 20 trials for testing).The results displayed are the averages

274   of 20 simulations for each combination of parameters. As expected, there was an overall decay of

275   performance when increasing the background firing rate and jitter, due to the increasing difficulty in

276   extracting time patterns. Still, the WI method provided the best decoding accuracy in nearly all cases,

277   except when very large jitters were used (of the order of, or larger than the time patterns themselves).

278   In this case, all information in the time patterns was destroyed and only spike count carried information.

279   Similar results were obtained for Examples 2 and 3 (Supporting Figure 1).

280   We then investigated how the number of trials used for training the decoder, and for selecting

281   the set of response features used for decoding, affected performance. To do so, we repeated the

282   analysis of Figure 4A but systematically varied the training set size (from 5 to 65 trials per stimulus in 5-

283   trial steps). We used 20 simulations for each method and set of parameters (jitter and background rate).

284   Figure 4B reports results for all 3 examples, averaging across all baseline firing and jitter values of Figure

285   4A. The performance of the WI, PCinfo and no-reduction method increased monotonically with the

286   number of trials used for training. In contrast to the PCinfo and no-reduction methods, the WI method

287   reached a value close to its maximum performance within less than 20 trials, stressing its robustness to

288   undersampling.

289   To evaluate the efficiency of each method in reducing the dimensionality of the responses

290   without loss of information, we computed the performance of the naïve Bayesian classifiers as a

291   function of the number of response features used for classification. For this, we ranked the features

292   either by variance, in the case of PCAvar; or by information, in the case of wavelets and PCAinfo.

293   Additionally, we performed a similar analysis by ranking the response time bins by their amount of

294   information and then selecting only the *n* most informative ones (referred hereafter as 'binned

295   responses'). Figure 5A reports the results for each of the three examples of Figure 3, when using a

296   training set of 15 trials per stimulus. While the WI method needed 10 or fewer response features to

297   reach maximal performance, all other methods needed a larger number of response features, which also

298   varied substantially across examples. Thus, the performance for these latter methods was very sensitive

299   to the number of response features used. In sum, the WI method reduced the dimensionality of

300   responses in a more efficient and robust way.

301   Interestingly, the performance using principal components with largest variance (PCVar) had a

302   much steeper increase of information with the number of features compared to the performance

303   obtained with PC coefficients chosen based on information (PCinfo). This seemingly counterintuitive

304   result can be attributed to the low number of trials (15 per stimulus) used for training in this case, which

305   gave a relatively poor estimation of information carried by each principal component. To verify this, we

306   ran the same analysis as in Figure 5A but using 50 trials per stimulus (Figure 5B). As expected, with the

307   larger training set PCInfo showed a clear increase in performance and was much more efficient than

308   PCVar for small numbers of features. Consistent with the results shown in Figure 4B, increasing the

309   number of trials had little impact on the performance of the WI method, thus highlighting its robustness

310   to sample size.

311

PERFORMANCE WITH DATA FROM THE MONKEY AUDITORY CORTEX

313   Single neuron recordings were performed in primary auditory (A1) cortex in response to a 40-

314   sec long sequence of pseudorandom tones (see Methods for details). We divided the sequence into 500-

315   ms time intervals and denoted each time interval as a different discrete "stimulus". For this dataset, we

316   trained naïve Bayesian decoders to predict which of different chunks of the time-varying stimulus was

317   being presented. In total, 34 responsive neurons (with >1 Hz mean rate) recorded in 12 sessions were

318   included in this analysis. Each session consisted of 50 to 60 presentations of the stimuli, which we

319   separated into two non-overlapping sets of training and test trials.

320   For each neuron, we first evaluated the performance of the various methods with a time

321   resolution of 1-ms, using 15 trials for training. Figure 6A shows the decoding performance of each

322   neuron using the different approaches described above (y-axis), against the performance achieved with

323   wavelets (x-axis). Note that the WI method outperformed the other methods for virtually all neurons.

324   Figure 6B shows decoding performance versus training set size and reveals that the WI method

325   performed significantly better than the other methods. Moreover, performance with wavelets

326   decreased only slightly when decreasing the training set size and was close to optimal with as few as 10-

327   15 trials. In contrast, the other methods showed a marked decrease when using few trials. Consistently,

328   however, all methods revealed that spike timing contained more stimulus information than the spike

329   count.

330   We then quantified the impact of temporal precision used to quantify the neural responses.

331   Figure 6C shows that decoding performance was maximal when using a bin size of about 5-10ms (and

332   more towards 10 ms for the PCA based methods). This result is comparable to the optimal resolution

333   previously reported for this data using a direct information estimate (Kayser et al. 2010). In all cases, the

334   performance decreased as the bin size increased, meaning that larger bin sizes missed relevant

335   information arising from the precise pattern of firing of these neurons. Moreover, the performance of all

336   methods also decreased for bin sizes smaller than 5-10 ms. This arises because the outcome of

337   increasing temporal precision is the trade-off of two opposing effects. One the one hand, a finer

338   resolution leads to a potentially higher information content in the neural responses. Because of the data

339   processing inequality (Quiroga and Panzeri 2009), increasing resolution can only increase or leave

340   invariant the information available in the responses. On the other hand, a finer resolution increases the

341   dimensionality of the responses, thus making it more difficult for a decoder to extract the available

342   information. A drop in decoding performance when increasing the resolution thus means that from that

343   resolution onward the additional information available at finer resolution is no more sufficient to

344   overcome the added difficulty in decoding many extra and weakly informative dimensions. Thus, the

345   ability to extract more information at finer resolutions with WI compared to other methods is due to the

346   optimal dimensionality reduction implemented in the WI method. For example, with a bin size of 1 ms

347   (i.e. an order of magnitude increase of the dimensionality of the response space used for decoding), the

348   performance with wavelet decreased only about 15% with respect to the 5-10 ms resolution, while

349   decreases of about 30% and more than 50% were observed for the no-reduction and the PCA-based

350   methods, respectively.

351   Next, we evaluated the ability of all methods to extract information from populations of

352   simultaneously recorded cells. For this, we used the data of 10 (out of 12) sessions where 2 or more

353   responsive neurons were recorded simultaneously. For each session, all possible combinations of a

354   varying number of cells were used. In this case, we did not set a minimum of wavelet coefficients or

355   principal components for each neuron; i.e., if a neuron had no significantly informative features then

356   this neuron would provide no features to the decoder. Results (Figure 6D) show that the decoding

357   performance with wavelets was significantly larger than the one achieved with the other methods.

358

359 *PERFORMANCE WITH DATA FROM THE RAT BARREL CORTEX*

360   To evaluate the potential of the WI method to optimally extract time localized information, we

361   analyzed neuronal responses in the rat barrel cortex. In this data, a precise onset time is given by the

362   time of whisker stimulation. We used a naïve Bayesian decoder, as before, to classify which vibrissa

363   (from the set of C1, C2, C3, D1, D2, D3, E1, E2 and E3) was stimulated in each trial. Forty-eight trials were

364   available for each vibrissa, and we randomly assigned half of these trials for training and the other half

365   for testing. Figure 7A shows the responses of a representative neuron. Responses were binned in 1-ms

366   windows, as before, and considered in a 200-ms response window starting at stimulus onset. Figure 7B

367   compares the WI method with the other approaches for 84 responsive neurons (with >1 Hz mean firing

368   rate). The WI method again outperformed all other methods for the large majority of the cells. The left

369   panel of Figure 7C displays the decoding performance as a function of the post-stimulus window (i.e. a

370   window of 200-ms means taking the whole response shown in Figure 7A). Consistent with the previous

371   results, the WI method gave the best results for nearly all time windows considered. In contrast to the

372   auditory data above (Figure 6), in this case the 'no reduction' approach did not give a good performance

373   because only a fraction of the bins provided relevant information. Moreover, while for wavelets the

374   performance kept increasing within increasing post-stimulus window, for the no-reduction and the PCA-

375   based methods the performance decayed or remained at the same level. This again indicates that these

376   methods do not perform as well as wavelets when the dimensionality of the response increases. We

377   also observe that for this data the 'spike count' decoding provided better results than both PCA-based

378   (and the no-reduction) approaches. However, this was in part due to systematic errors in the decoder.

379   For example, the PCAvar decoder classified D1 responses as D1, D2 or D3 (but still ruling out the

380   possibility that other vibrissae were stimulated). In order to verify this, we computed the information

381   between the predicted and the actual stimuli from the confusion matrices, as shown in the right panel of

382   Figure 7C (see Methods for details). In this case, PCAvar provided more information than the spike

383   counts. As before, in nearly all cases the wavelet-based decoding provided the best results.

384   Figure 7D shows both the decoding performance (left) and the information extracted from the

385   confusion matrices (right) as a function of the bin size used. The peak of performance and information

386   was at a bin size of about 25 ms for all methods, except wavelets. In particular, all methods showed

387   decay in information for larger bin sizes because with larger bin sizes the information given by precise

388   time patterns is lost. But interestingly, all methods except wavelets showed also decay in performance

389   (and information) for bin sizes smaller than 25 ms. As with the monkey data, this is due to the increase

390   in the response space dimensionality accompanying the increase in resolution. With this same data, a

391   previous work (Panzeri et al. 2001) reported an optimal bin size smaller than 25 ms, using a method

392   analogous to the 'no-reduction' shown here, i.e. calculating the mutual information from the binned

393   responses. But in that case, a high dimensional space was avoided by considering a much smaller

394   response window of 20 ms. In this regard, the advantage of wavelets is crucial whenever the optimal

395   response window is not known *a priori*.

396   Finally, we studied the information carried by populations of neurons. For this, we assumed that

397   all neurons were recorded simultaneously (an approach that doesn't take into consideration the effects

398   of correlations) and repeated the procedure used in Figure 6C. Since 100 neurons were available, we

399   averaged across 30 randomly chosen combinations for each number of neurons. Additionally, to avoid

400   an excessive number of features we only allowed a maximum of 5 wavelet coefficients or principal

401   components per neuron (instead of 25, as before). Results are shown in Figure 7E. The performance

402   achieved with wavelets surpassed the one achieved with the other methods. Except for the no-

403   reduction and PCAinfo cases (due to the abovementioned limitation in dealing with high dimensional

404   response spaces), the performance increased monotonically with the number of neurons.

405

406   *COMPARISON WITH THE METRIC SPACE METHOD*

407   Next we compared the WI method with the MS approach for different q values (see Material

408   and Methods), both with the simulated and real data reported above. Figure 8A shows the results of

409   such comparison with same 20 simulations illustrated in Figure 2A, using different background firing

410   rates. Note that WI is more robust to increases of noise levels for this example. Figure 8B shows the

411   results for the simulations shown in Figure 4A (to save space, results with 32 and 64-ms jitter are not

412   shown since in these cases the performance with all methods was close to chance), and Figure 8B the

413   result for the 3 examples presented in Figure 4B. Altogether, we observe an overall better performance

414   with the WI method proposed here. Note also that the performance of the MS method is dependent of

415   the choice of the parameter q. For instance, a q value of 0.128 gave the best performance with a jitter of

416   16 ms but also the worst performance with a jitter of 2ms. Conversely, a q value of 0.362 gave the best

417   performance for the 2-ms jitter case but a relative poor performance for larger jitters. In this respect,

418   the advantage of the WI method is that it doesn't require the tuning of any parameter and automatically

419   gives a performance that in most cases surpassed the one obtained with the MS method, even when

420   choosing the optimal q value.

421   Figure 8C displays results as function of training set size (as in Figure 4B). For these examples, a

422   q value of 0.128 gave the best overall results for the MS method. These results were, however, not as

423   good as the ones obtained with the WI method. We note that the MS method was remarkably robust to

424   undersampling, as can be seen in the results of Example 2, where it outperformed wavelets when less

425   than 10 trials per stimulus was used for training, likely due to the fact that this low number of trials was

426   not sufficient for a good selection of wavelet coefficients.

427   Results of the comparison between the WI and MS method for the real data are shown in Figure

428   9. In particular, Figure 9A (left panels) displays the results obtained for the A1 monkey neurons (as in

429   Figure 6) using four representative q values for the MS method. For this dataset, the average

430   performance obtained with the WI method was significantly better than that obtained with the MS

431   method for all q values (paired t-test, p<0.05 in all cases; see right panel in Figure 9A). Figure 9B shows

432   the results obtained for the barrel cortex neurons. As with the monkey data, performance with the WI

433   method was significantly better than that obtained with the MS approach for all q values (paired t-test,

434   p<10^{-23}in all cases; see right panel in Figure 9B). The relatively poor performance of the MS method for

435   the rat dataset is likely due to the compact time localization of the informative spikes.

436

437   *DENOISING SPIKE PATTERNS*

438   We next investigated whether the WI method could be used to denoise single-trial spike trains.

439   Denoising entails, in brief, identifying spikes correlated with the informative wavelet coefficients and

440   discarding the remaining, non-informative ones (see Methods). Figure 10Ashows the denoising of the

441   time pattern of Example 3 (Figure 3) with a jitter of 4-ms jitters and background firing rate of 64-Hz. We

442   observe that a large amount of the background non-informative spikes were removed and the

443   remaining spike rasters after denoising (Figure10A, bottom left) were very similar to the spike patterns

444   embedded in the data (see Figure 3). A similar reduction of background 'noisy' activity is evident when

445   comparing the PSTHs before and after denoising (Figure10A, right Panels).

446   To quantify these observations, we repeated this procedure varying systematically the jitter and

447   background firing rate (using 15 trials per stimulus). Figure 10B shows the number of errors obtained by

448   the denoising procedure for different jitters as a function of background rate. Here, we defined *errors* as

449   the sum of false positives (not deleting a spike corresponding to background activity) and false negatives

450   (incorrectly discarding a spike that belonged to the informative time pattern). For comparison, we also

451   calculated the number of errors obtained when thresholding the original PSTHs but without a prior

452   wavelet denoising (green traces). This was done to assess whether denoising could be achieved by a

453   simple PSTH thresholding. Further, to show that results were not just due to the smaller number of

454   spikes obtained after denoising, we also calculated the number of errors obtained when randomly

455   erasing the same number of spikes (black lines). In general, the wavelet-based denoising approach gave

456   the lowest number of errors. Results obtained for the other examples of Figure 3 were similar and are

457   shown in Supporting Figure 2.

458   Next, we applied this denoising approach to the spikes of a representative neuron taken from

459   the auditory dataset. Figure 11A shows the raster plots before (top) and after (bottom) denoising, where

460   it is clear that time patterns (spikes consistent across trials for each stimulus) are more easily visualized

461   after denoising. Given this encouraging result, we then asked whether wavelet denoising could lead to

462   obtaining cleaner and sharper spike-triggered averages (STA) of the stimulus. STAs are commonly used

463   representations to assess which stimulus features (out of the many in a complex dynamic stimulus) drive

464   the neurons' responses (Dayan and Abbot 2005). For this, we used a time-frequency representation of

465   the acoustic stimulus (see Methods), and we computed the STA in the form of the Spectro-Temporal

466   Receptive Field, i.e. the average frequency spectrum of the stimulus around the time of spiking. To test

467   the effectiveness of denoising in removing the detrimental effect of non-stimulus driven spikes, we first

468   computed the STA for the original response, then added background Poisson noise with a mean firing of

469   100Hz and computed the STA in this noisy condition. Finally, we denoised the data with the Poisson

21

470   noise and recomputed the STA. Results are shown in Figure 11B. The original STA showed a clear tuning

471   to stimuli with energy in the 200-300 Hz frequency range, approximately 50 ms preceding the spike

472   occurrences. This stimulus selectivity was dramatically diluted after adding the background noise, but it

473   was again recovered after denoising.

474

475   *INFORMATION IN THE SYNCHRONOUS FIRING OF NEURONS*

476   We next investigated whether the WI method could be extended to quantify the information

477   conveyed by the synchronous firing of pairs of neurons, by assuming that in this case the informative

478   wavelet coefficients for both neurons should covary across trials (see Methods). For comparison, we

479   also evaluated the information carried by: 1) coincident spikes – i.e. spikes from different neurons fired

480   within a short time window (Grun et al. 2002) – although it should be noted that, in contrast to WI, with

481   this method a single, unique, time scale that defines "coincidence" has to be defined *a-priori*; 2) an

482   implementation of the MS approach, calculated by computing the distance between the spike trains for

483   the pair of neurons as a function of q, and then using a decoder similar to the one used for wavelets

484   (using the MS distance instead of the wavelet distance). For this analysis, we used the 3 q values shown

485   in Figure 8.

486   Performance was tested by simulating the activity of a pair of neurons with a correlated firing

487   during a 'synchrony window' (200-360 ms in the case of the example displayed in Figure 12A, left) during

488   the presentation of one of two stimuli (stimulus 1). In the rest of the response window we added spikes

489   generated independently, following a Poisson distribution with a mean firing probability equal to the

490   one of the synchronous spike pattern (40-Hz). Therefore, by construction, all the information about the

491   stimuli was only given by the transient synchronization.

22

492   We ran several simulations like the one illustrated in Figure 12A (left panel), varying

493   systematically the duration of the synchrony window and adding time jitters in the co-activation. We

494   used 20 trials per stimulus for training the decoders and 20 for testing. Figure 12B shows the decoding

495   performances for different jitters as function of the length of the synchrony duration. Clearly, the WI

496   method outperformed the coincidence count for nearly all window sizes and jitters. The overall

497   decoding performance for all methods is shown in Figure 12C.

498   We then extended the spike train denoising procedure described above to better visualize

499   informative patterns of synchronous firing. For this, we used the distances of the selected wavelets. The

500   mask for a given stimulus (analogous to the denoised PSTHs in the previous case) was constructed by

501   averaging the distances of each wavelet coefficient across trials and then adding together the mean

502   distance of each coefficient multiplied by its time support (with a value of 1 within the time range

503   spanned by the wavelet and 0 elsewhere). Figure 12A (right panels) displays the raster plots of the

504   neurons shown on the left panels after denoising, where it is clear that the coincident activations are

505   highlighted and non-informative spikes deleted.

506

507   **DISCUSSION**

508   A key problem in assessing the contribution of precise time patterns to sensory coding is the

509   high dimensionality of the datasets. While several statistical methods have been developed to correct

510   for sampling biases arising in these cases (Montemurro et al. 2007; Nemenman et al. 2004; Paninski

511   2003; Panzeri et al. 2007; Panzeri and Treves 1996; Strong et al. 1998), these methods are still of limited

512   value when dealing with long response spaces with high temporal precision. Other approaches to tackle

513   this issue have been proposed, for example, based on simplifying the structure of interactions among

514   possible information-carrying symbols with a minimum information loss (Ganmor et al. 2011; Panzeri

515   and Schultz 2001; Shew et al. 2011),developing binless estimations (Victor 2002) or defining spike train

516   distances to quantify information (Victor and Purpura 1996).

517   In a classic work, Richmond and colleagues proposed to reduce the dimensionality of the

518   response space by using PCA (Richmond and Optican 1987). However, PCA offers no time resolution -

519   thus being limited for characterizing time-resolved patterns - and it relies on identifying directions of

520   maximum variance of the data, which may or may not match the dimensions with relevant information.

521   Here, following the general idea of dealing with high dimensional spaces by implementing a

522   dimensionality reduction that captures relevant information, we proposed a new computational

523   approach to assess information carried by time patterns in single and multiple neurons. This approach is

524   based on: i) extracting features of the spike trains with the wavelet transform, ii) a dimensionality

525   reduction by which a subset of wavelet coefficients are selected using information theory, and iii) a

526   quantification of time-pattern information by using a decoding approach.

527   With both simulated and real data we demonstrated a robust performance of the WI method in

528   capturing meaningful information in the spike trains, without committing to specific assumptions about

529   the time scales at which information is encoded and even capturing information in localized patterns at

530   multiple time scales. This feature is of utmost importance considering that recent studies have shown

531   that neural responses carry complementary information at a number of different time scales, ranging

532   from millisecond precise spike patterns to slow rate variations or slow network oscillations on the scale

533   of hundreds of milliseconds (Bullock 1997; Fairhall et al. 2001; Lisman 2005; Nadasdy 2009; Panzeri et al.

534   2010; Victor 2000). However, it has been challenging to characterize how these neural responses work

535   together to represent information, because most spike train analysis methods are committed from the

536   beginning by the specific choice of an optimal time scale.

537   The performance of the WI method was minimally affected by increases in the dimensionality of

538   the responses (obtained by increasing the time resolution, the length of the response considered, or the

539   number of neurons) compared to other methods. In fact, compared to methods like PCA, using the

540   whole (binned) response space (i.e. without any dimensionality reduction), or using spike counts the WI

541   method was able to extract more information from the spike trains, as quantified by decoding

542   performance. In addition, the information obtained with wavelets was more robust to varying degrees

543   of background activity and jitter in the precise timing of the spikes. For the cortical data analyzed here,

544   high information values were found with PCA, but only when the optimal time scale (i.e. response length

545   and resolution bin) was considered. Crucially, information values obtained with wavelets were much less

546   sensitive to the choice of the time scale used to study the neural responses. In our view, these

547   advantages arise because the convolution with Haar functions implemented with the dyadic wavelet

548   transform identifies local contrasts at different time locations and at different time scales. In other

549   words, the wavelet transform offers a time resolved, multi-scale representation that automatically and

550   efficiently represents time patterns of different lengths and resolutions, appearing at different times.

551   The advantages of wavelets for spike train analysis reported here are in line with the reported

552   advantages of wavelets for spike sorting (Quiroga 2012; Quiroga et al. 2004) and for denoising evoked

553   potentials (Ahmadi and Quian Quiroga 2013).

554   We also compared the WI method to the MS method (Victor and Purpura 1996), a widely used

555   metric based approach to estimate information in spike trains. We found that overall the WI method

556   performed better more robustly than the MS method. In particular, MS results varied substantially

557   depending on the choice of q (defining the weight given to precise timing vs. number of spikes) and no

558   q-value gave good results in all conditions – i.e. for different firing rates and jitters. So, it may not be

559   possible to find a single q-value that is suitable for different neurons in a dataset (with different firing

560   rates, time pattern resolution, degrees of information, etc). On the contrary, WI is parameter free and

561   performed well in all tested cases. With the real data, the performance of the MS method was lower

562   than WI, likely because of the compact time localization of the informative spikes, something that is

563   captured by the WI method when doing the selection of informative wavelet coefficients. Another

564   aspect of practical importance is that the MS method took significantly longer to compute than WI. For

565   the monkey data, while the WI method required consistently between 100 and 150 seconds to compute

566   the results for each neuron, the MS method processing time was strongly correlated with the mean rate

567   of the responses: computations for neurons firing around 2.5 Hz on average took about 200 seconds or

568   more per q value, whereas computations for neurons firing around 5 Hz took around 400 seconds. Since,

569   as in previous works (Roussin et al. 2012; Victor and Purpura 1996), results are calculated for about

570   dozens of different q values, computations with the MS method took about two orders of magnitude

571   longer compared to wavelets.

572   A key strength of the WI approach is its data-robustness. In particular, the WI method required

573   fewer trials to achieve optimal results, and needed fewer features to represent the relevant information

574   in the spike trains compared to other methods. This efficient compression of the responses mitigates

575   the 'curse of dimensionality' and allows the analysis of larger responses, the use of higher resolutions,

576   and also the possibility of population coding analyses where features of several neurons are considered

577   together. This represents a significant advance compared to the other dimensionality reduction based

578   approaches examined in our study, which in some cases tended to underestimate the time resolution or

579   the amount of information in the spike trains because actual increases in information were

580   counterbalanced by the limitations of these methods to deal with higher dimensional responses.

581   It is of interest to briefly discuss how the data robustness of the WI method (due to its highly

582   efficient reduction of dimensionality) relates to the biases in extracting information from neural

583   responses often discussed in the literature (Panzeri et al. 2007). First, the success of the WI method in

584   effectively compressing the responses to a very small number of informative dimensions leads to a

585   strong reduction in the *downward bias* in decoding performance, which is given by a limited number of

586   data for training the decoder (Jacobs et al. 2009; Quiroga and Panzeri 2009). This is demonstrated by the

587   success of the WI method to extract high information values even for very fine temporal resolutions in

588   real data and the nearly optimal performance obtained with relatively few trials with the simulated

589   data. As a rule of thumb, 10 or more trials per stimulus were found to be enough to avoid a major

590   downward bias problem (see Figure 4b). Second, the higher information values achieved by the WI

591   method allow reducing (and correcting more efficiently) the *upward bias* in estimation of information

592   from the confusion matrix due to the limited number of experimentally available test data. As shown in

593   Panzeri and Treves (1996), this upward bias is roughly proportional to the number of different stimuli

594   that are predicted by the decoder when a given stimulus is presented. As a rule of thumb, the

595   corrections for the upward bias in the confusion matrix information calculation work well if the number

596   of different stimuli that are predicted by the decoder when a given stimulus is presented is 2 to 4 times

597   smaller than the number of trials per stimulus (Panzeri et al. 2007). The WI method gives less decoding

598   errors, and therefore a smaller upward bias in confusion matrix information calculations.

599   We stress that with the WI method we do not estimate directly the mutual information

600   contained in the stimulus response probabilities, either in the form of binary words, as in (Strong et al.

601   1998), or in the form of PC scores, as in (Optican and Richmond 1987). Instead, we limited ourselves to

602   the calculation of cross-validated confusion matrices. These estimations give a lower bound to the

603   information available in the data (Quiroga and Panzeri 2009). However, it is in principle possible to

604   extend the WI method to help also direct calculations of information from neural activity (Strong et al.

605   1998). It is well known that such direct computations of information from binary response words are in

606   practice possible only for short response windows, due to the curse of dimensionality (Kayser et al.

607   2009). In this respect, the WI method may be further developed to extend the applicability of direct

608   calculations of information to longer windows by providing a relatively low-dimensional representation

609   of the binary word response, thus providing more data robust direct calculations of information. Such

610   implementations are beyond the scope of the present paper but are ripe for future work.

611   We have also shown that the WI approach can be adapted to denoise spike trains and to

612   estimate and visualize correlations across neurons. In particular, we found an improved performance in

613   estimating information in the correlated firing of neurons compared to a MS approach (Victor and

614   Purpura 1996) or the standard technique of analyzing coincidence counts, what has been called "unitary

615   events" (Grun et al. 2002). As in the case of other methods to assess time-pattern information, the

616   caveat of the unitary event approach is that it is highly dependent on the time resolution used to bin the

617   spike trains. For instance, co-activations can be missed if the time scale at which they occur does not

618   match the window used for binning the spiking activity (Lopes-dos-Santos et al. 2013). Likewise, results

619   with the MS approach rely on the choice of an optimal q-value. The advantage of wavelets in this

620   respect is the fact that it allows evaluating correlated firing at different time scales and for specific time

621   localizations.

622   Finally, the possibility of denoising spike trains allows a much clearer visualization of informative

623   time pattern and a better characterization of the neuron's selectivity (though reverse correlation), as we

624   illustrated with data from monkey auditory cortex. The spike train denoising based on WI could in

625   principle be used for a number of other novel applications. For example, it can be used to better

626   characterize relationships between spiking activity and local field potentials (particularly their phase at

627   selected frequency bands) carrying out information about the stimuli. The better visualization of the

628   stimulus driven spikes can be also useful to study relationships between stimulus-driven and "internal

629   state" components of neural activity, a topic of current interest in systems neuroscience (Harris and

630   Thiele 2011). In general, the more accurate characterization of the neurons' tuning offered by WI

631 denoising will likely offer important practical advantages for the discovery of the "thesaurus" that

632 relates sensory stimulus to neural responses, and could further facilitate the understanding of what

633 features encoded by the neurons do contribute to behavior.

634   A MATLAB implementation of the WI is available from: www.le.ac.uk/csn/WI.

635

636

637 **ACKNOWLEDGEMENTS**

643

644

FIGURE LEGENDS

646   **Figure 1. Description of the method.** (**A**) Simulated neural responses to 2 stimuli. (**B**) Responses are

647   binned and convolved with Haar wavelets. Note that for the example in (A), the time pattern associated

648   with Stimulus 1 matches both scale and time localization of the wavelet in the first level displayed in

649   blue; whereas the time pattern associated with Stimulus 2 has a larger jitter and fits with a wavelet in a

650   higher scale (shown in red). (**C**) Wavelet coefficients with significant time pattern information are

651   identified. Dashed lines represent statistical threshold for information significance for each

652   decomposition level (see Methods). The red and blue circles denote the information values

653   corresponding to the wavelet coefficients in B. (**D**) Decoding results in the form of confusion matrices

654   using the total spike count (left) or the informative wavelet coefficients (right) as inputs to the decoder.

655   **Figure 2. Decoding time patterns at different time scales.** (**A**) Simulated responses to 4 stimuli. Marks

656   represent spikes at a given time (horizontal axis) in a given trial (vertical axis). Twenty trials per stimulus

657   are displayed (dashed lines separate trials of different stimuli) and only the first 152 milliseconds of each

658   trial are displayed for clearer visualization. Vertical axes denote the size of the bins used (light gray:1-ms

659   and dark gray:8-ms bins).(**B**) Confusion matrices of naïve Bayesian decoders trained with PCs extracted

660   from responses binned with 1 or 8-ms bins (left and middle panel, respectively), and based on

661   informative wavelet coefficients (right panel). Decoding performances are shown in white. (**C**) Confusion

662   matrices obtained when using wavelet coefficients from specific decomposition levels, as labeled.

663   **Figure 3. Results for different examples, background firing and jitter.** The panels show 3 examples of

664   simulated responses to 4 different stimuli. Ten trials per stimulus are displayed. The background firing

665   was modeled as a Poisson process with mean rate of 4 (left panels) or 16 spikes/second (right panels).

666   Stimulus-specific time patterns were added to the background firing with a 2- (left panels) or 8-ms (right

667   panels) jitter.

668 **Figure 4. Performance of the WI method.** (**A**) The different plots show the time pattern information,

669 quantified in terms of decoding performance using spike counts (black), principal components with

670 largest variance (purple), principal components with significant information (blue), no dimensionality

671 reduction (green), and wavelets (red). Each data point represents the average of 20 simulations (error

672 bars show SEM) for Example 1 of Figure 2 with varying jitters and firing rates, as labeled.(**B**) Relative

673 decoding performance as a function of the training set size for the different methods. Results are

674 averages across all combinations of background firing rates and jitters. Note that in nearly all cases the

675 wavelet-based algorithm outperformed the other methods.

676 **Figure 5. Information with varying number of features.** (**A**) Decoding performance for all 3 examples,

677 with 8-Hz background firing rate and 4-ms jitter, as function of number of features used for decoding

678 (mean ± SEM; 40 simulations). For all these simulations we chose intermediate values of background

679 firing rate (8-Hz) and jitter (4 ms). Features were ranked by information (wavelets, binned responses and

680 PCinfo) or by variance (PCvar). The training set size was of 15 simulations per stimulus. (**B**) Same as A,

681 but using with a training set of 50 trials per stimulus.

682 **Figure 6. Time patterns in responses from monkey A1 neurons.** (**A**) Time pattern information of 34 A1

683 neurons extracted using the PCA-based and no-reduction methods (y-axis), against the performance

684 achieved with wavelet (x-axis). Most cases are below the diagonal (dashed lines), indicating a better

685 performance of the wavelet-based method. The training set size was of fifteen trials per stimulus. (**B**)

686 Average decoding performance for different methods as a function of the training set size. Bars denote

687 standard error of the mean. The performance with wavelets was larger than with the other methods. (**C**)

688 Average decoding performance for different methods as a function of the bin size (15 trials per stimulus

689 used for training). (**D**) Average decoding performance for the different methods as a function of number

690 of simultaneously recorded neurons used. Marks show averages across 10 sessions and bars denote

691   standard error of the mean. As in (A), fifteen trials per stimulus were used for training. The performance

692   with wavelets was significantly larger when considering more than 1 neuron.

693   **Figure 7. Time patterns in rat barrel cortex neurons.** (**A**) Responses of a representative neuron to

694   mechanical stimulation of different whiskers (C1-3, D1-3, E1-3). Each dot denotes a spike in a given trial

695   (vertical axis) at a given time (horizontal axis). Forty-eight trials were recorded for each whisker. Stimuli

696   were delivered at 0-ms and lasted 100-ms, which corresponded to the OFF response. (**B**) Decoding

697   performance for 84 responsive neurons with wavelets (x-axis) compared to spike count, PCAinfo, PCAvar

698   and with no dimensionality reduction, respectively. Most entries are below the diagonal (dashed line),

699   thus indicating a better performance with wavelets. (**C**) Left: Mean decoding performance for different

700   methods as a function of the post-stimulus time window. Right: Mean information for the different

701   methods, estimated from the confusion matrices (see text for details). Solid lines denote averages and

702   shaded areas SEM. (**D**) Left: Decoding performance for different methods as function of bin size. Right:

703   Stimulus information for different methods estimated from confusion matrices as function of bin size.

704   (**E**) Mean decoding performance for the different methods as a function of number of neurons used.

705   Solid lines show averages across a set of 30 randomly chosen neurons (except for 100 neurons, where

706   only 1 combination is possible) and shaded areas denote SEM (barely visible due to the very small error).

707   As before, note the better performance of wavelets for nearly all post-stimulus time windows and any

708   number of neurons considered.

709   **Figure 8. Comparison of MI and MS methods with simulated data.** (**A**) Performance of WI and MS (for 3

710   different values of q, as labeled) methods for multi-scale responses of Figure 2A with varying

711   background rates. Shown are mean and SEM across 20 simulations for each noise level. (**B**) Performance

712   of WI and MS for the same examples in Figure 4A. For space reasons, we do not show results with 32

713   and 64-ms jitters (performance was close to chance levels in all these cases). (**C**) Same as Figure 4B

714   comparing the WI and MS approaches.

715   **Figure 9. Comparison of the MI and MS methods with real data.** (**A**) Left panels: Same as in Figure 6A

716   for the WI and the MS method with 4 representative q values. In average, decoding performance

717   achieved with wavelets was 26 ± 8%, 24 ± 8%, 144 ± 35% and 449 ± 73% (mean ± SEM) above the one

718   obtained by the MS method when q employed was 0.045, 0.128, 0.362 and 1.024, respectively. Right

719   panel: Decoding performances for WI and MS methods (for the same q values). Shown are mean and

720   SEM. Wavelets gave better results for all q values (n = 34, paired t-test, p<0.05). (**B**) Left panels:

721   comparison of WI and MS for the same data as in Figure 7B. For this dataset, decoding performance

722   achieved with wavelets was 46 ± 5%, 54 ± 5%, 61 ± 5% and 68 ± 6% (mean ± SEM) superior the one

723   obtained by the MS method (same q values as in A). Right panel: Comparison of WI and MS performance

724   for the rat data. Wavelets gave better results for all q values (n = 84, paired t-test, p<10^{-23}).

725   **Figure 10. Denoising time patterns with simulated data.** (**A**) Left panels: A 100-trial realization (25 trials

726   per stimulus) of Example 3 from Figure 2, before (top) and after (bottom) denoising. Right panels:

727   Peristimulus-Time-Histograms (PSTHs) for each of the 4 stimuli before (top) and after denoising

728   (bottom). (**B**) Number of errors (see text) after denoising (in red) obtained for Example 3with different

729   jitters (same values used in Figure 3) as function of background firing rate. Results are averages across

730   40 simulations (error bars denote SEM). For comparison, results obtained when thresholding the original

731   PSTHs (without prior wavelet denoising; see text for details) are shown in green. Results obtained by

732   randomly erasing the same number of spikes as with wavelet denoising are displayed in black. In most

733   cases, the lowest number of errors was obtained after wavelet denoising.

734   **Figure 11. Denoising time patterns with monkey A1 responses.** (**A**) Responses to 8 stimuli of a

735   representative A1 neuron before (top) and after (bottom) denoising. Gray lines separate trials

736 corresponding to the different stimuli.(**B**) Spike trigger averages (STA) of the time-frequency

737 representation of the acoustic stimuli computed from the original spike trains (top), when adding a

738 Poisson noise with 100-Hz mean firing rate (middle), and after denoising the spike train with the Poisson

739 noise (bottom). Note that adding the Poisson noise dilutes the original time frequency representation,

740 which is recovered after denoising.

741 **Figure 12. Extracting information from spike synchrony.** (**A**) Left: Responses of two simulated neurons

742 to two stimuli. Spikes between 200 and 360-ms were coincident in trials corresponding to stimulus 1 (in

743 red). This transitory synchrony was the only source of stimulus information in the responses. Right:

744 Denoised spike trains. (**B**)Performance obtained with the wavelet-based method, by the MS approach

745 and with the number of spike coincidences for different spike jitters and synchrony durations. (**C**)

746 Average decoding performance for all methods across all jitter and background levels.

747

748

749

750

REFERENCES

Adrian ED, and Zotterman Y. The impulses produced by sensory nerve-endings. Part II. The response of a Single End-Organ. *Journal of Physiology* 61: 151-171, 1926.

Ahmadi M, and Quian Quiroga R. Automatic denoising of single-trial evoked potentials. *NeuroImage* 66: 672-680, 2013.

Arabzadeh E, Panzeri S, and Diamond ME. Deciphering the spike train of a sensory neuron: Counts and temporal patterns in the rat whisker pathway. *Journal of Neuroscience* 26: 9216-9226, 2006.

Bullock TH. Signals and signs in the nervous system: The dynamic anatomy of electrical activity is probably information-rich. *Proceedings of the National Academy of Sciences of the United States of America* 94: 1-6, 1997.

Dayan P, and Abbot L. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. The MIT Press, 2005.

de Ruyter van Steveninck RR, Lewen GD, Strong SP, Koberle R, and Bialek W. Reproducibility and variability in neural spike trains. *Science* 275: 1805-1808, 1997.

Di Lorenzo PM, Chen J-Y, and Victor JD. Quality Time: Representation of a Multidimensional Sensory Domain through Temporal Coding. *Journal of Neuroscience* 29: 9227-9238, 2009.

Eckhorn R, and Popel B. Rigorous and extended application of information theory to afferent visual system of the cat. II. Experimental results. *Biological Cybernetics* 17: 7-17, 1975.

Fairhall AL, Lewen GD, Bialek W, and van Steveninck RRD. Efficiency and ambiguity in an adaptive neural code. *Nature* 412: 787-792, 2001.

Fisher RA. The use of multiple measurements in taxonomic problems. *Annals of Human Genetics* 7: 10, 1936.

Foffani G, Morales-Botello ML, and Aguilar J. Spike Timing, Spike Count, and Temporal Information for the Discrimination of Tactile Stimuli in the Rat Ventrobasal Complex. *Journal of Neuroscience* 29: 5964-5973, 2009.

Fontanini A, and Katz DB. State-dependent modulation of time-varying gustatory responses. *Journal of Neurophysiology* 96: 3183-3193, 2006.

Fotowat H, Harrison RR, and Gabbiani F. Multiplexing of motor information in the discharge of a collision detecting neuron during escape behaviors. *Neuron* 69: 147-158, 2011.

Ganmor E, Segev R, and Schneidman E. The architecture of functional interaction networks in the retina. *J Neurosci* 31: 3044-3054, 2011.

Grun S, Diesmann M, and Aertsen A. Unitary events in multiple single-neuron spiking activity: 1. Detection and significance. *Neural Computation* 14: 43-80, 2002.

Harris KD, and Thiele A. Cortical state and attention. *Nat Rev Neurosci* 12: 509-523, 2011.

Harvey MA, Saal HP, Dammann JF, III, and Bensmaia SJ. Multiplexing Stimulus Information through Rate and Temporal Codes in Primate Somatosensory Cortex. *Plos Biology* 11: 2013.

Jacobs AL, Fridman G, Douglas RM, Alam NM, Latham PE, Prusky GT, and Nirenberg S. Ruling out and ruling in neural codes. *Proceedings of the National Academy of Sciences of the United States of America* 106: 5936-5941, 2009.

Kayser C, Logothetis NK, and Panzeri S. Millisecond encoding precision of auditory cortex neurons. *Proceedings of the National Academy of Sciences of the United States of America* 107: 16976-16981, 2010.

Kayser C, Montemurro M, Logothetis N, and Panzeri S. Spike-Phase Coding Boosts and Stabilizes Information Carried by Spatial and Temporal Spike Patterns. *Neuron* 61: 597-608, 2009.

Laurent G, Wehr M, and Davidowitz H. Temporal representations of odors in an olfactory network. *Journal of Neuroscience* 16: 3837-3847, 1996.

797 **Lebedev MA, Mirabella G, Erchova I, and Diamond ME**. Experience-dependent plasticity of rat barrel
798 cortex: Redistribution of activity across barrel-columns. *Cerebral Cortex* 10: 23-31, 2000.
799 **Lisman J**. The theta/gamma discrete phase code occuring during the hippocampal phase precession may
800 be a more general brain coding scheme. *Hippocampus* 15: 913-922, 2005.
801 **Lopes-dos-Santos V, Ribeiro S, and Tort ABL**. Detecting cell assemblies in large neuronal populations.
802 *Journal of Neuroscience Methods* 220: 149-166, 2013.
803 **Magri C, Whittingstall K, Singh V, Logothetis NK, and Panzeri S**. A toolbox for the fast information
804 analysis of multiple-site LFP, EEG and spike train recordings. *Bmc Neuroscience* 10: 2009.
805 **Mallat S**. *A Wavelet Tour of Signal Processing: The Sparse Way*. Academic Press, 2008.
806 **Montemurro MA, Panzeri S, Maravall M, Alenda A, Bale MR, Brambilla M, and Petersen RS**. Role of
807 precise spike timing in coding of dynamic vibrissa stimuli in somatosensory thalamus. *Journal of
808 Neurophysiology* 98: 1871-1882, 2007.
809 **Nadasdy Z**. Information encoding and reconstruction from the phase of action potentials. *Front Syst
810 Neurosci* 3: 6, 2009.
811 **Nemenman I, Bialek W, and van Steveninck R**. Entropy and information in neural spike trains: Progress
812 on the sampling problem. *Physical Review E* 69: 2004.
813 **Optican LM, and Richmond BJ**. Temporal encoding of two-dimensional patterns by single units in
814 primate inferior temporal cortex. III. Information theoretic analysis. *Journal of Neurophysiology* 57: 162-
815 178, 1987.
816 **Paninski L**. Estimation of entropy and mutual information. *Neural Computation* 15: 1191-1253, 2003.
817 **Panzeri S, Brunel N, Logothetis NK, and Kayser C**. Sensory neural codes using multiplexed temporal
818 scales. *Trends in Neurosciences* 33: 111-120, 2010.
819 **Panzeri S, Petersen RS, Schultz SR, Lebedev M, and Diamond ME**. The role of spike timing in the coding
820 of stimulus location in rat somatosensory cortex. *Neuron* 29: 769-777, 2001.
821 **Panzeri S, and Schultz SR**. A unified approach to the study of temporal, correlational, and rate coding.
822 *Neural Comput* 13: 1311-1349, 2001.
823 **Panzeri S, Senatore R, Montemurro M, and Petersen R**. Correcting for the sampling bias problem in
824 spike train information measures. *Journal of Neurophysiology* 98: 1064-1072, 2007.
825 **Panzeri S, and Treves A**. Analytical estimates of limited sampling biases in different information
826 measures. *Network-Computation in Neural Systems* 7: 87-107, 1996.
827 **Quiroga RQ**. Spike sorting. *Current Biology* 22: R45-R46, 2012.
828 **Quiroga RQ, Nadasdy Z, and Ben-Shaul Y**. Unsupervised spike detection and sorting with wavelets and
829 superparamagnetic clustering. *Neural Computation* 16: 1661-1687, 2004.
830 **Quiroga RQ, and Panzeri S**. Extracting information from neuronal populations: information theory and
831 decoding approaches. *Nature Reviews Neuroscience* 10: 173-185, 2009.
832 **Quiroga RQ, Reddy L, Koch C, and Fried I**. Decoding visual inputs from multiple neurons in the human
833 temporal lobe. *J Neurophysiol* 98: 1997-2007, 2007.
834 **Richmond BJ, and Optican LM**. Temporal encoding of two-dimensional patterns by single units in
835 primate inferior temporal cortex. II. Quantification of response waveform. *Journal of Neurophysiology*
836 57: 147-161, 1987.
837 **Rieke F, Warland D, van Steveninck RdR, and Bialek W**. *Spikes: Exploring the Neural Code*. A Bradford
838 Book, 1999, p. 416.
839 **Roussin AT, D'Agostino AE, Fooden AM, Victor JD, and Di Lorenzo PM**. Taste coding in the nucleus of
840 the solitary tract of the awake, freely licking rat. *J Neurosci* 32: 10494-10506, 2012.
841 **Shadlen MN, and Newsome WT**. Noise, neural codes and cortical organization. *Current Opinion in
842 Neurobiology* 4: 569-579, 1994.
843 **Shannon C**. A mathematical theory of communication. *Bell System Technical Journal* 27: 379-423 and
844 623 - 656, 1948.

845   **Shew WL, Yang H, Yu S, Roy R, and Plenz D**. Information capacity and transmission are maximized in
846   balanced cortical networks with neuronal avalanches. *J Neurosci* 31: 55-63, 2011.
847   **Strang G, and Nguyen T**. *Wavelets and Filter Banks*. Wellesley-Cambridge Press, 1996, p. 500.
848   **Strong SP, Koberle R, van Steveninck RRD, and Bialek W**. Entropy and information in neural spike trains.
849   *Physical Review Letters* 80: 197-200, 1998.
850   **Victor JD**. Binless strategies for estimation of information from neural data. *Phys Rev E Stat Nonlin Soft*
851   *Matter Phys* 66: 051903, 2002.
852   **Victor JD**. How the brain uses time to represent and process visual information. *Brain Res* 886: 33-46,
853   2000.
854   **Victor JD, and Purpura KP**. Nature and precision of temporal coding in visual cortex: A metric-space
855   analysis. *Journal of Neurophysiology* 76: 1310-1326, 1996.

856