

1 **Precise hyperacuity estimation of spike timing from calcium** 2 **imaging**

3

4 Huu Hoang^{1*}, Masa-aki Sato^{1*}, Shigeru Shinomoto^{1,2}, Shinichiro Tsutsumi³, Miki
5 Hashizume⁴, Tomoe Ishikawa⁵, Masanobu Kano³, Yuji Ikegaya⁵, Kazuo Kitamura⁶,
6 Mitsuo Kawato¹, Keisuke Toyama^{1#}

7

8 ¹ATR Brain Information Communication Research Laboratory Group, Advanced
9 Telecommunications Research Institute International, Kyoto, Japan

10 ²Department of Physics, Kyoto University, Kyoto, Japan

11 ³Department of Neurophysiology, Graduate School of Medicine, The University of Tokyo,
12 Tokyo, Japan

13 ⁴Department of Biochemistry, Faculty of Medicine, Saitama Medical University

14 ⁵Graduate School of Pharmaceutical Sciences, The University of Tokyo, Tokyo, Japan

15 ⁶Department of Neurophysiology, Faculty of Medicine, University of Yamanashi,
16 Yamanashi, Japan

17

18

19

20

21 * Equally contributing first authors

22 # Correspondence: Keisuke Toyama, toyama@atr.jp

23

24

25 **Summary**

26 Two-photon imaging is a major recording technique in neuroscience, but it suffers
27 from several limitations, including a low sampling rate, the nonlinearity of calcium
28 responses, the slow dynamics of calcium dyes and a low signal-to-noise ratio, all of which
29 impose a severe limitation on the application of two-photon imaging in elucidating
30 neuronal dynamics with high temporal resolution. Here, we developed a hyperacuity
31 algorithm (HA_time) based on an approach combining a generative model and machine
32 learning to improve spike detection and the precision of spike time inference. First,
33 Bayesian inference estimates the calcium spike model by assuming the constancy of the
34 spike shape and size. A support vector machine employs this information and detects
35 spikes with higher temporal precision than the sampling rate. Compared with conventional
36 thresholding, HA_time improved the precision of spike time estimation up to 20-fold for
37 simulated calcium data. Furthermore, the benchmark analysis of experimental data from
38 different brain regions and simulation of a broader range of experimental conditions
39 showed that our algorithm was among the best in a class of hyperacuity algorithms. We
40 encourage experimenters to use the proposed algorithm to precisely estimate hyperacuity
41 spike times from two-photon imaging.

42 Introduction

43 Recently, two-photon imaging has been one of the major means of recording
44 multineuronal activities in neuroscience to obtain the precise morphology and location of
45 the target neurons because of its high spatial resolution¹⁻⁶. However, its utility is still
46 constrained by its relatively low temporal resolution due to the mechanical scanning of
47 two-photon rays. The other problems are the nonlinearity, the slow dynamics and the low
48 signal-to-noise ratio (SNR) of the calcium (Ca) responses⁷⁻¹⁰. Many algorithms have been
49 proposed to reconstruct spike trains from Ca imaging data, including conventional
50 thresholding¹¹, deconvolution¹²⁻¹⁵, template matching¹⁶⁻²⁰, Bayes inference²¹⁻²³ and
51 machine learning^{24,25}, to overcome these problems. Few of them, however, have
52 addressed the two challenging goals simultaneously: reliable spike detection and spike
53 time estimation with high temporal precision in the presence of the nonlinearity, slow
54 dynamics and low SNR of the Ca responses²⁶. For the former goal, the spike dynamics
55 of the target neurons and/or kinematics of the Ca responses may vary dramatically across
56 brain regions and different Ca dyes. For the latter goal, a trade-off between the number
57 of recorded neurons and temporal resolution exists. The slow kinematics and the low
58 SNR of the currently available Ca dyes may also limit the temporal precision of the
59 information conveyed by the Ca responses. These factors impair reliable spike detection
60 as well as precise spike time estimation for high-frequency firing that is frequently
61 encountered in cortical cells²⁷⁻²⁹.

62 Here, we propose an approach combining a generative model of Ca responses
63 including nonlinearity and dye dynamics with a supervised classifier to overcome the
64 aforementioned difficulties. Our hyperacuity algorithm, named HA_time (HyperAcuity time
65 estimation), estimates the Ca spike model by Bayesian inference assuming size and
66 shape constancy, compensates for the nonlinearity of the Ca responses, and detects
67 spikes from Ca imaging data by a support vector machine (SVM) using the ground-truths,
68 i.e., simultaneously recorded electrical spikes, as supervised information. To achieve
69 hyperacuity precision, spike timings were calibrated to minimize the residual errors in
70 model prediction using the hyperacuity vernier scale. This approach benefits from the
71 advantages of both generative models and supervised learning. On the one hand, the Ca

72 spike model is utilized to provide supplemental information for spike detection as well as
73 to estimate the spike times at higher temporal precision than the sampling resolution. On
74 the other hand, the supervised learning compensates for fluctuations in the Ca responses
75 due to noise and sampling jitters, which are not considered by the generative model. As
76 a consequence, HA_time can improve both the spike detection and spike time estimation
77 of two-photon recordings.

78 A simulation study confirmed that compared with the thresholding algorithm,
79 HA_time improved the temporal precision by 2-20-fold. The previous algorithms have
80 aimed to improve spike detection as well as spike time estimation with higher temporal
81 precision than that expected for the sampling rate of two-photon recordings. They
82 assumed generative models for spike generation and maximized the likelihood of the
83 estimates^{17,21,22}. Accordingly, hyperacuity performance was limited to only the cases
84 where the Ca responses satisfy the assumptions of generative models. To prove the
85 advantages of the approach combining the generative model and supervised learning,
86 we compared our method with four previous hyperacuity algorithms^{16,17,22,23}. The
87 benchmark results for the experimental data sets showed that HA_time was among the
88 best across three brain regions: the cerebellar, hippocampal and visual cortices.
89 Furthermore, the simulation analysis conducted across a broad range of parameters for
90 the experimental conditions, including the mean neuronal firing frequency, the
91 nonlinearity, the decay time of the Ca dyes, and the sampling rate, providing useful
92 information for users to select the most suitable algorithms for the given experimental
93 conditions and highlighting the advantages of our algorithm over the other ones under the
94 high firing frequency and/or strong nonlinearity conditions frequently encountered in the
95 cortical cells of behaving animals.

96

97 **Results**

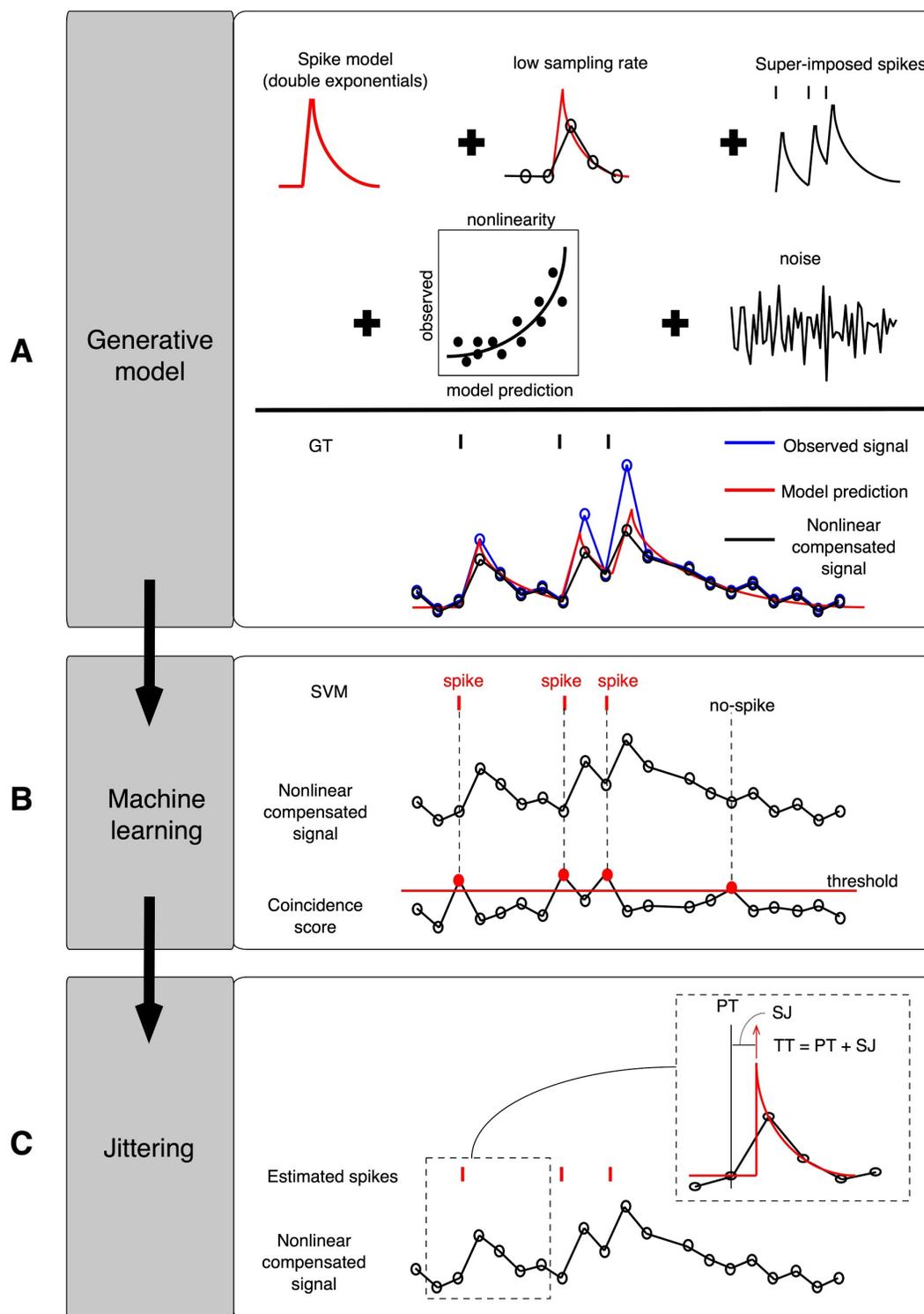
98 **Hyperacuity algorithm for spike timing estimation**

99 Our hyperacuity algorithm, HA_time, was conducted in three steps: 1) Bayesian
100 inference of the Ca spike model from Ca imaging data, 2) spike detection by SVM assisted

101 by the matching information of the Ca imaging data with the Ca spike model and 3)
102 hyperacuity spike time estimation to minimize the errors between the Ca response model
103 prediction and the recorded Ca imaging data using the hyperacuity vernier scales. We
104 noted here that the term “Ca spike model” indicates the constant Ca transient (i.e., the
105 amplitude and shape) of a single spike, whereas “Ca response model” is a generative
106 model of the Ca spike model, superposition of multiple spikes, nonlinearity of the Ca
107 responses and noise (see below).

108 We assumed that the Ca imaging data were sampled from the Ca spike model
109 (double exponentials) with variable sampling jitters between the onsets of the Ca spike
110 and the sampling times. The Ca spike responses were first linearly superimposed for
111 multiple spikes in short intervals and added with the Gaussian noise. The sub- or
112 superlinearity of the Ca responses was determined by comparing the observed Ca
113 imaging data with the data predicted by the Ca response model. We compensated for the
114 nonlinearity by inversely transforming the observed Ca imaging data by nonlinearity
115 models fitted by logarithmic functions (Fig. 1A).

116 Next, we estimated the coincidence score as a convolution of the first-order
117 derivative of the Ca imaging data and that of the Ca spike model. A coincidence score
118 threshold was used to sample the data segments as spike candidates. Here, the threshold
119 and segment size were optimized to maximize the F1-score of the training data (see
120 Methods). The SVM was trained to classify the sampled data segments into spike or non-
121 spike segments. For this purpose, we fed the sampled Ca imaging data and the
122 coincidence scores as the primary and attribute inputs, respectively, to the SVM and used
123 the electrical spikes (ground truth) as the teaching signals (Fig. 1B).



124

125 **Figure 1: The hyperacuity support vector machine (HA_time) algorithm.** A: The generative model
 126 assumed that Ca responses were sampled from the Ca spike model (double exponentials) with variable
 127 sampling jitters (SJs) due to the low sampling rate, superimposed by multiple spikes, fluctuated by
 128 nonlinearity and supplemented with Gaussian noise. Nonlinearity observed in the data was compensated

129 by the nonlinearity model defined by logarithmic functions of the observed Ca imaging data and Ca
130 response model prediction. B: The coincidence score, computed by convoluting the first derivative of Ca
131 imaging data and that of the Ca spike model, was used to select spike candidates. Support vector machine
132 (SVM) was trained to classify spikes and non-spikes from spike candidates with the Ca imaging data, the
133 coincidence scores and the electrical spikes as feature, attribute, and teaching signals, respectively. C: The
134 true spike time (TT) was estimated as the sum of SJ and pseudo-spike time (PT, the point that exceeds the
135 threshold) minimizing the residual error of the Ca response model prediction using hyperacuity vernier 10-
136 fold finer than the sampling interval.

137

138 For the test data, spike candidates were sampled in the same way as the training
139 data, and the trained SVM detected the spikes among the candidates. We tentatively
140 determined the time when the coincidence score exceeded the threshold as the pseudo-
141 spike time (PT in Fig. 1C inset) and estimated the sampling jitters (SJs) to minimize the
142 errors between the Ca response model prediction and the Ca imaging data using the
143 hyperacuity vernier scales (10-fold finer time bin than the sampling interval). The true
144 spike time (TT) was calculated as the sum of the PT and SJ (Fig. 1C).

145

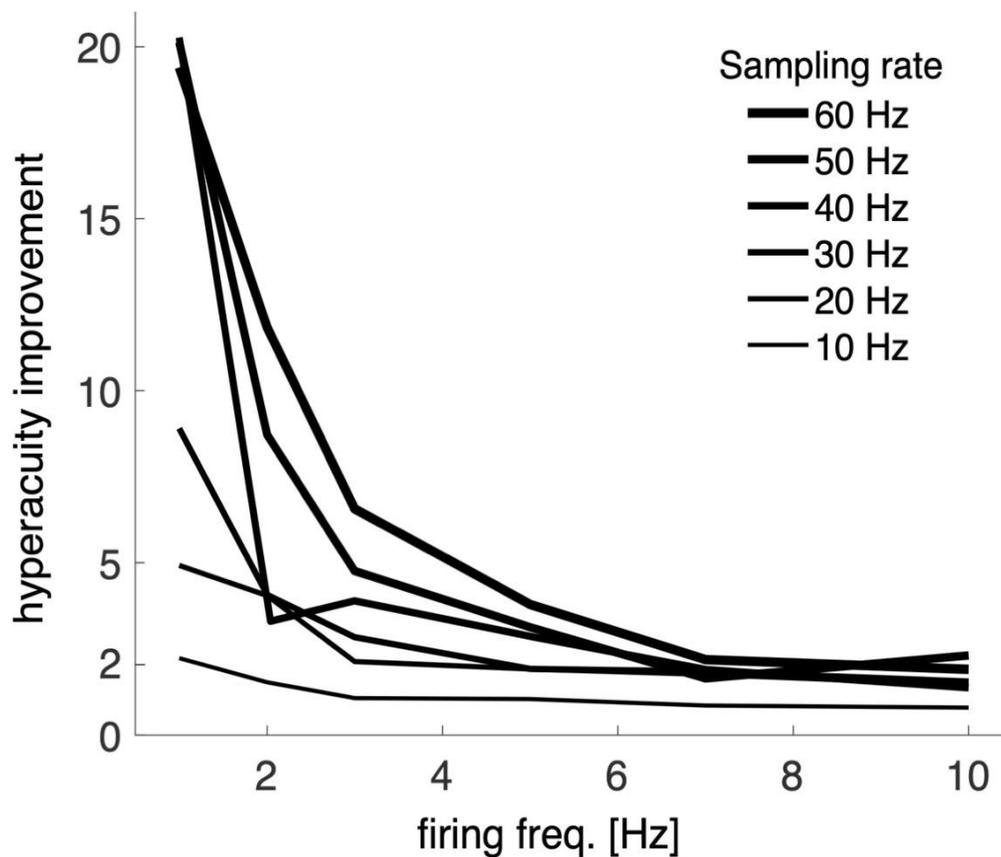
146 **Hyperacuity improvement of HA_time in simulation**

147 To illustrate the hyperacuity improvement of HA_time, we compared our method
148 with the thresholding algorithm using the simulated data with no nonlinearity and fast Ca
149 decay time (see Methods). Here, the threshold of the thresholding algorithm was
150 optimized to maximize the F1-score of the training data.

151 Figure 2 shows the hyperacuity improvement, determined as the ratio of the
152 mean of the spike time errors for the traditional thresholding algorithm to that for HA_time
153 (see Methods), as a function of the mean firing frequency of the simulated spike train.
154 Compared with the thresholding algorithm, in cases with a high sampling rate of 40-60
155 Hz, HA_time improved the temporal precision more than 5-fold under low firing rate
156 conditions (<5 Hz) and maintained an approximately 2-fold improvement under high firing
157 rate conditions, probably due to the decreased performance of both algorithms (thicker
158 lines, Fig. 2). Such a tendency was also found in the cases with a lower sampling rate of

159 10-30 Hz, although the hyperacuity improvement was smaller (thinner lines). These
160 results indicated that even under fairly simple data conditions, for which the conventional
161 methods are widely used, HA_time was able to provide a 2-20-fold improvement in the
162 temporal precision compared to that provided by the thresholding algorithm. For example,
163 if we take an ideal case of a 60 Hz sampling frequency and 2 Hz firing frequency, the
164 hyperacuity improvement was 12, meaning that the effective sampling rate is
165 approximately $60 \times 12 = 720$ Hz, which is quite satisfactory.

166



167

168 **Figure 2: The hyperacuity improvement of HA_time.** The hyperacuity improvement of HA_time
169 compared with conventional thresholding as a function of the firing frequency at various sampling rates of
170 10-60 Hz (encoded by line thickness). The nonlinearity parameter α was fixed at 1, and τ_2 and SNR were
171 fixed at 0.2 s and 5, respectively.

172

173 **Application of HA_time to experimental data**

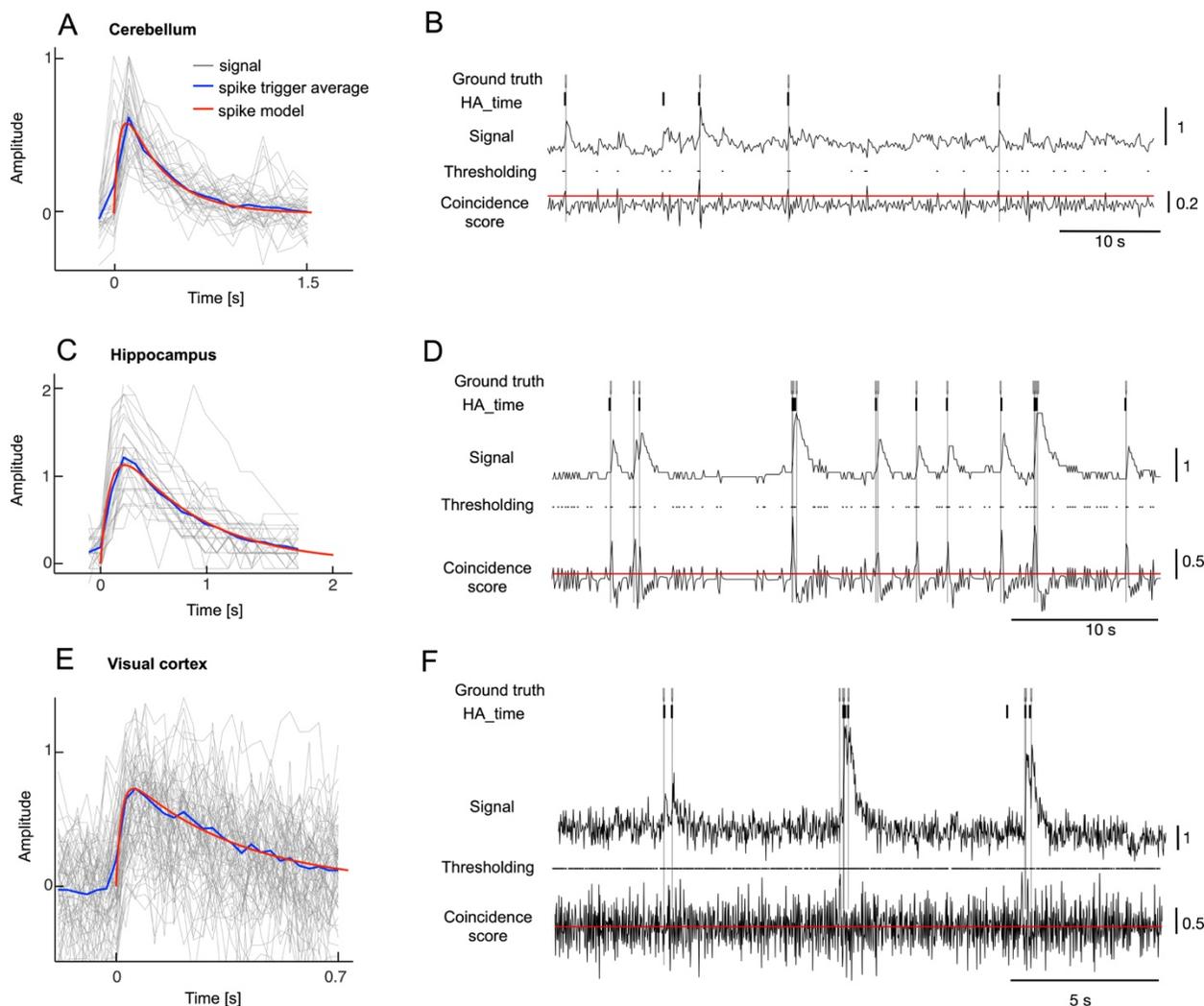
174 We applied HA_time to noisy Ca imaging data obtained in cerebellar, hippocampal
175 and primary visual cortical cells by two-photon recording with relatively low sampling
176 rates.

177 For the two-photon recording of the Ca response from five Purkinje cells in the
178 cerebellum (the dye, Cal-520, and sampling rate, 7.8 Hz), we sampled thirty-six data
179 segments (segmental length, 2 s), each including a single electrical spike from the
180 simultaneous electrical recording (sampling rate, 20 KHz), and constructed the Ca spike
181 model by Bayesian inference. In agreement with the assumption of spike-shape
182 constancy, the Ca spike model ($\tau_1 = 0.05$ s, $\tau_2 = 0.4$ s, red trace) was slightly faster in the
183 rise time (red trace of Fig. 3A) than the electrical spike-triggered average of the Ca
184 imaging data, while the amplitude of the Ca spike model roughly agreed with those for
185 the spike-triggered averages. The longer time course of the spike-triggered response may
186 be due to the sampling jitters. We also conducted a Bayesian estimation of the Ca spike
187 models for the data from the entire hippocampus (n=9 cells, sensor, OGB1-AM and
188 sampling rate, 10 Hz) and visual cortex (n=11 cells, sensor, GCAMP6f, sampling rate of
189 60 Hz). We avoided segments that contain burst activity (interspike interval < 2 s) since
190 the Ca responses in the hippocampus and visual cortex showed strong nonlinearity during
191 the burst activity (see below). A similar tendency to that in the cerebellum data was also
192 noticed in the Ca spike model (red traces, Figs. 3C&E), which was faster in the rise time
193 than the spike-triggered averages (blue traces) for both the hippocampus and visual
194 cortex data. The dynamics of the dyes estimated by our Bayesian method were in
195 agreement with those reported in previous studies ($\tau_1 = 0.01$ and $\tau_2 = 0.2$ s for GCAMP6f³⁰
196 and $\tau_1 = 0.1$ and $\tau_2 = 0.75$ s for OGB1-AM¹¹).

197 Figure 3B illustrates the performance of HA_time in detecting spikes from the Ca
198 imaging data of cerebellar cortex cells. The coincidence thresholding (black dots in Fig.
199 3B) detected the true spikes (ground truth, gray bars) as well as many false-positive
200 spikes. HA_time effectively selected the true spikes, rejecting many false alarm spikes
201 from the spike candidates. Comparison of the spikes detected by HA_time (dark bars)

202 with the ground truths (gray bars) indicated that HA_time almost perfectly detected spikes
203 and correctly estimated the spike time from the Ca imaging data for the cerebellum (Fig.
204 3B). HA_time also performed almost completely correctly, rejecting many false alarms
205 detected by conventional thresholding, and it estimated the spike times for the
206 hippocampus and visual cortex data (Figs. 3D and F).

207



208

209 **Figure 3: Estimation of the Ca spike model and spike detection by HA_time.** A: Ca spike model (red
210 trace, τ_1 and τ_2 , 0.05 and 0.4 s) and spike-trigger averaged Ca responses synchronized with the onsets of
211 electrical spikes. The Ca spike model and the spike-triggered average were estimated for the thirty-six
212 electrical spikes of five Purkinje cells. Ordinate, amplitude of Ca responses normalized for the peak of the
213 maximum Ca imaging data for the individual cells. Abscissa, time after the onset of the electrical spikes. B:

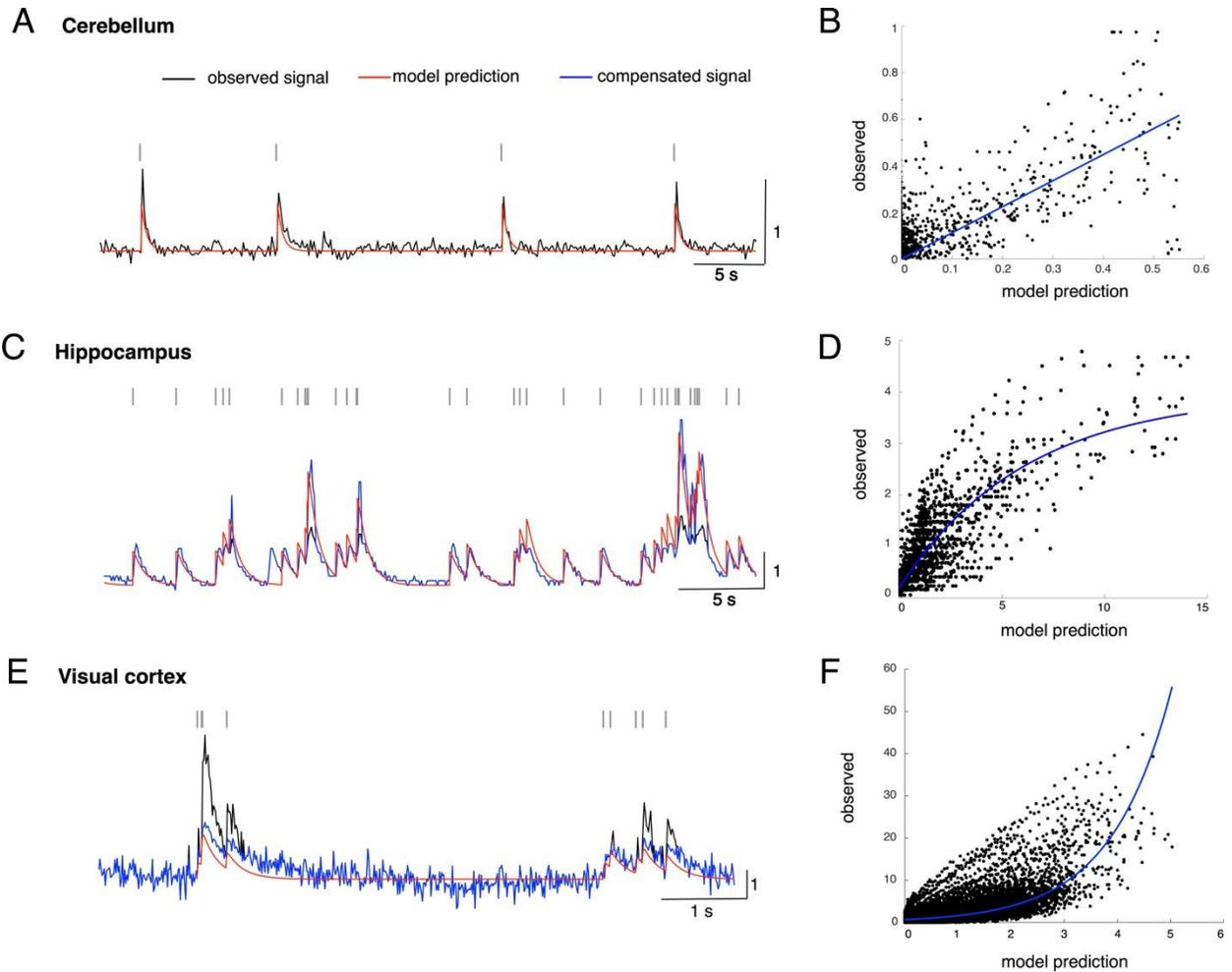
214 Spike detection by HA_time. The top and bottom traces represent Ca imaging data and the coincidence
215 score of the first-order differential of the Ca imaging data with that of the Ca response model. The candidate
216 spikes detected by conventional thresholding, those estimated by HA_time and the electrical spikes (ground
217 truth) are denoted by black dots, thick and thin bars, respectively. Red lines indicate the thresholds for
218 conventional thresholding. C, D and E, F: similar to A and B but for the hippocampus (τ_1 and τ_2 , 0.1 and
219 0.75 s) and visual cortex data (0.01 and 0.2 s), respectively.

220

221 **Nonlinearity analysis of the experimental data sets**

222 We found strong nonlinearity in the Ca imaging data of the hippocampal and visual
223 cortex data during burst activities. Therefore, nonlinearity analysis was conducted by
224 plotting the amplitudes of the Ca imaging data as a function of the linear prediction of the
225 Ca response model for the entire cerebellum, hippocampus and visual cortex data.

226 The Ca imaging data of the cerebellum roughly agreed with the linear prediction
227 for spike trains (dark and red traces in Fig. 4A), and correspondingly, the regression
228 analysis revealed a fine match between the two (blue line in Fig. 4B, $y = 1.1 x$).
229 Conversely, the nonlinearity analysis of the Ca imaging data revealed significant sub- and
230 superlinearity in the hippocampus and visual cortex, respectively (dark and red traces in
231 Figs. 4C and E). The nonlinearity models were constructed by fitting the plots with
232 logarithmic functions (blue lines in Figs. 4D and F, $y = -3.8 e^{(-0.16x)} + 3.9$ and $y = 0.67$
233 $e^{0.88x}$). The nonlinearity in the hippocampus and visual cortex data was compensated by
234 multiplying the Ca imaging data by the inverse of the nonlinearity models (blue traces in
235 Figs. 4C and E). The compensated Ca imaging data were then fed into HA_time.



236

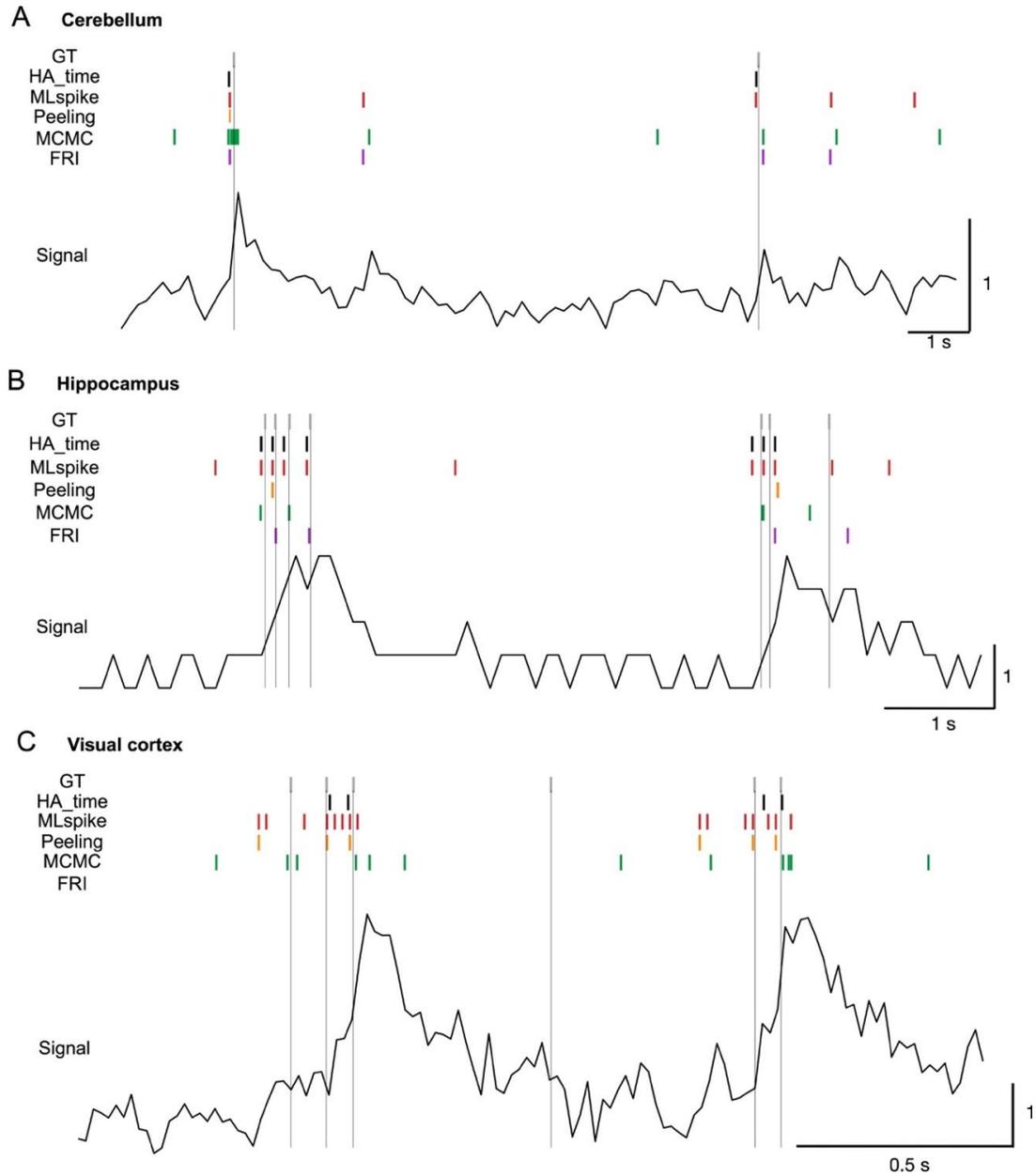
237 **Figure 4: Nonlinearity analysis of Ca imaging data.** A, C and E depict the Ca imaging data of the
238 cerebellum, hippocampus and visual cortex data, respectively. Black, red and blue traces represent the
239 observed Ca imaging data, linear prediction of the Ca response model for spike trains and compensated
240 Ca imaging data, respectively. B, D and F depict scatter diagrams for the Ca imaging data in the three
241 experimental data sets as a function of the linear prediction of the Ca response model for spike trains.

242

243 **Performance evaluation for experimental data**

244 The performance of HA_time in detecting spikes and in estimating the spike time
245 was studied for the cerebellum, hippocampus and visual cortex data by leave-one-out
246 cross-validation and compared with those of four benchmark algorithms^{16,17,22,23} (cf.
247 Methods).

248



249

250 **Figure 5: Spike detection by HA_time and benchmark algorithms.** A, B and C depict examples of spike
251 detection by HA_time (black bars), MLspike²³ (red), Peeling¹⁶ (orange), MCMC²² (green) and FRI¹⁷ (purple)
252 algorithms for the cerebellum, hippocampus and visual cortex data, respectively. Dark traces represent the
253 Ca responses. Thin vertical lines indicate the timing of the ground truth (GT) given by electrical spikes.

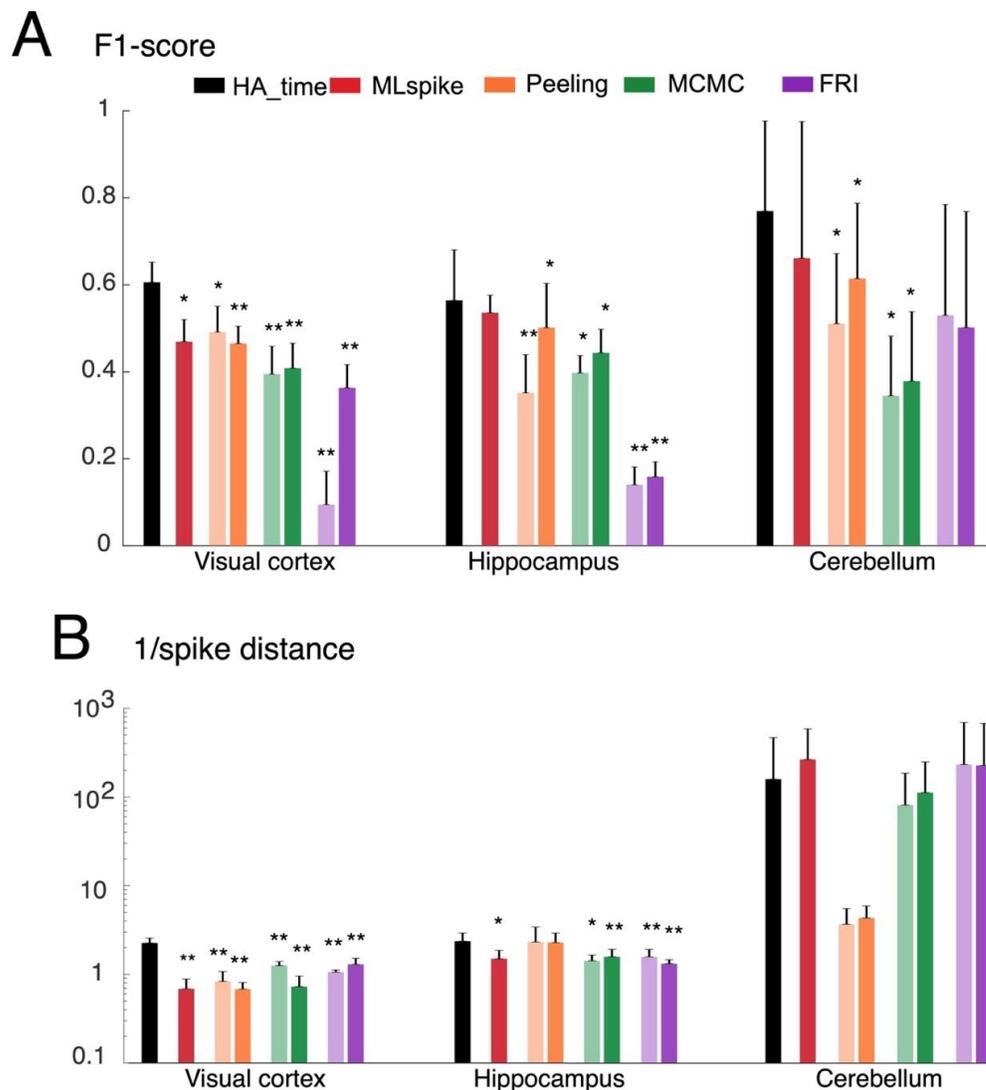
254

255 The spikes detected by HA_time (dark bars) matched fairly well with the ground
256 truths (gray bars) for all of the three experimental data sets. MLspike (red bars) performed
257 fine on the hippocampus data but rather poor on the cerebellum and visual cortex data
258 with many false positives. The remaining three algorithms (yellow, green and purple bars)
259 performed rather poorly for all of the three experimental data sets with many false-positive
260 or missing spikes (Figs. 5A-C).

261 We estimated spike detection performance by the F1-score of receiver operating
262 characteristic (ROC) analysis (cf. Methods). Among all of the algorithms, HA_time
263 performed best, with statistical significance in the F1-score for the visual cortex data (0.6
264 ± 0.04 for HA_time; 0.47 ± 0.05 for MLspike, $p = 0.01$ for HA_time vs. MLspike; $0.49 \pm$
265 0.06 for Peeling, $p = 0.01$ for HA_time vs. Peeling; 0.39 ± 0.06 for MCMC, $p = 0.002$ for
266 HA_time vs. MCMC; 0.09 ± 0.07 for FRI, $p = 0.002$ for HA_time vs. FRI). For the
267 hippocampus data, the superiority of HA_time (0.56 ± 0.11) was also clear, with
268 statistically significant F1-scores compared to those for the benchmark algorithms (0.35
269 ± 0.08 for Peeling, $p = 0.004$ for HA_time vs. Peeling; 0.39 ± 0.04 for MCMC, $p = 0.02$ for
270 HA_time vs. MCMC; 0.14 ± 0.04 for FRI, $p = 0.006$ for HA_time vs. FRI) except for
271 MLspike (0.53 ± 0.04). However, the significant superiority of HA_time (0.77 ± 0.21) over
272 the benchmark algorithms was limited to Peeling (0.51 ± 0.16 , $p = 0.02$ for HA_time vs.
273 Peeling) and MCMC (0.34 ± 0.13 , $p = 0.02$ for HA_time vs. MCMC) for the cerebellum
274 data. There was no statistical significance in the difference between HA_time and
275 MLspike (0.66 ± 0.31) or FRI (0.53 ± 0.25), probably due to the smaller number of cells
276 ($n=5$) in the cerebellum data (Fig. 6A).

277 The superiority of HA_time in terms of the precision of spike time estimation was
278 also found by the inverse of the spike distance³¹ (cf. Methods). Among all the algorithms,
279 HA_time performed best with statistical significance over all the benchmark algorithms
280 for the visual cortex data (1/spike distance, 2.2 ± 0.3 for HA_time; 0.7 ± 0.2 for MLspike,
281 $p = 0.002$ for HA_time vs. MLspike; 0.8 ± 0.2 for Peeling, $p = 0.002$ for HA_time vs.
282 Peeling; 1.2 ± 0.1 for MCMC, $p = 0.002$ for HA_time vs. MCMC; 1 ± 0.1 for FRI, $p = 0.002$
283 for HA_time vs. FRI). For the hippocampus data, except for Peeling (2.3 ± 1.1), HA_time
284 (2.3 ± 0.5) outperformed the benchmark algorithms with statistical significance (1.5 ± 0.3

285 for MLspike, $p = 0.02$ for HA_time vs. MLspike; 1.4 ± 0.2 for MCMC, $p = 0.01$ for HA_time
 286 vs. MCMC; 1.5 ± 0.3 for FRI, $p = 0.02$ for HA_time vs. FRI). However, for the cerebellum
 287 data, no statistical significance was found for HA_time compared with the benchmark
 288 algorithms (158.9 ± 309.9 for HA_time; 265.8 ± 325.1 for MLspike; 3.6 ± 1.8 for Peeling;
 289 81.4 ± 104.7 for MCMC; 233 ± 462.1 for FRI, cf. Fig. 6B).



290

291 **Figure 6: Performance benchmark for experimental data.** A and B, F1-score and inverse of spike
 292 distance for HA_time (black columns), MLspike (red), Peeling algorithm (orange), MCMC (green) and FRI
 293 (purple). The scores of the Peeling, MCMC and FRI algorithms for the original settings and those
 294 supplemented with the information provided by the HA_time are shown by dense and faint colors,
 295 respectively. The ordinates in A and B are shown in linear and log scale, respectively. The columns

296 represent the mean with error bars of +2SEM. Asterisks indicate significance level by Wilcoxon signed-rank
297 tests between HA_time and the benchmark algorithms. * and ** denote $p < 0.05$ and $p < 0.01$, respectively.

298

299 Notably, Peeling, MCMC and FRI were outperformed by HA_time and MLspike,
300 probably because they did not provide an effective routine to precisely estimate the
301 algorithm parameters in the presence of the nonlinearity of Ca responses, which is clearly
302 observed in the hippocampus and visual cortex data. In support of this view, their
303 performance was improved when we provided the parameters of the Ca spike model
304 estimated by HA_time and compensated for the nonlinearity of the Ca imaging data (cf.
305 light and dark bars, Fig. 6A-B).

306

307 **Performance evaluation for simulation data**

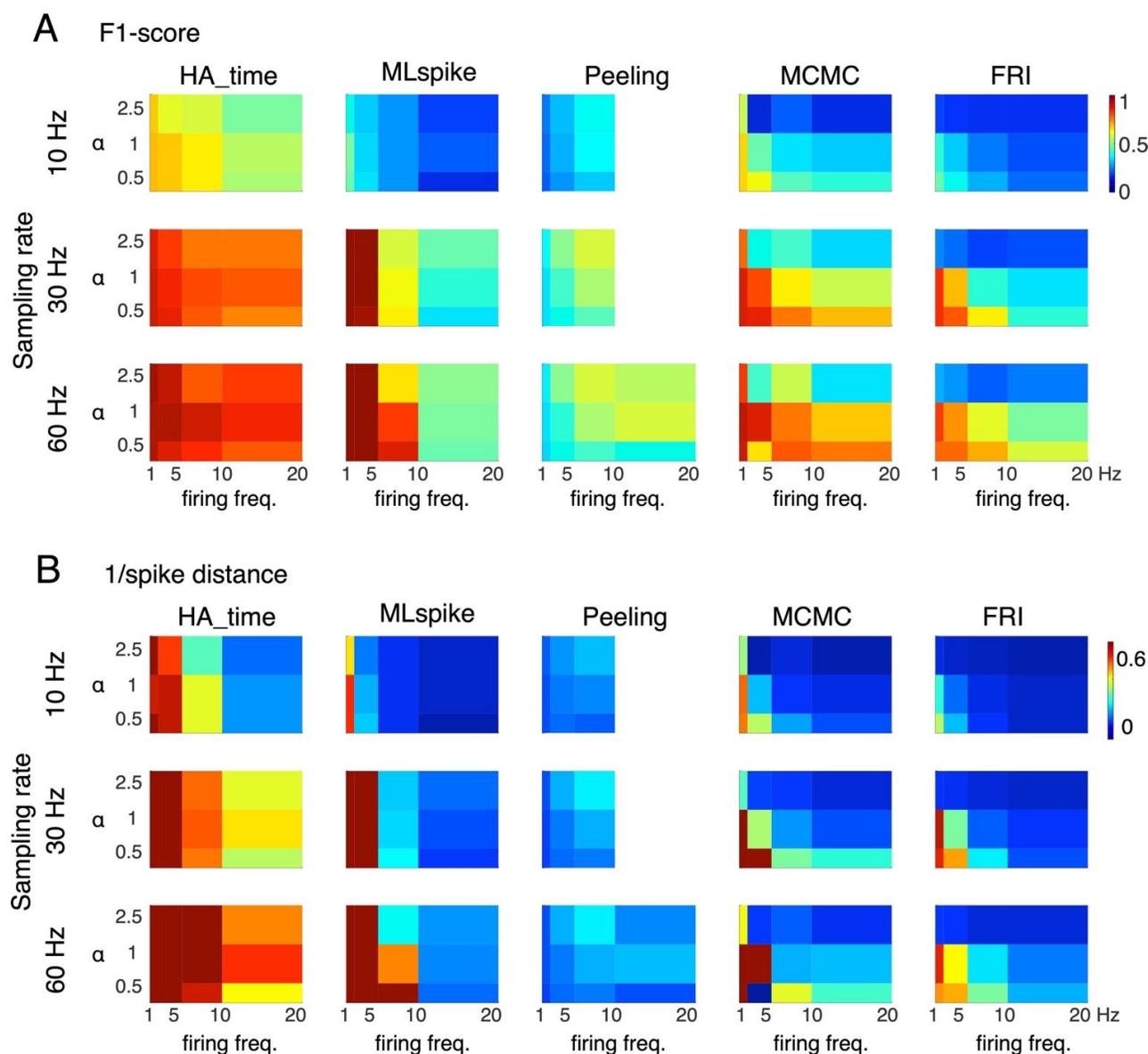
308 We further investigated the performance of HA_time and the benchmark
309 algorithms by simulating the Ca responses sampled for a broader range of conditions
310 than the experimental ones, including the mean firing frequency of the spike train, the
311 nonlinearity of the Ca responses, the sampling rate of the two-photon recording, the dye
312 dynamics of the Ca responses (time decay constant for the Ca responses) and the SNR
313 (cf. Methods).

314 The systematic analysis of performance as a function of the experimental
315 parameters revealed that the three parameters, the mean firing frequency, the
316 nonlinearity and the sampling rate, strongly influenced the relative performances of the
317 examined algorithms. In contrast, the remaining two parameters, the dye dynamics and
318 the SNR, altered the performance only in a quantitative manner without significant change
319 in the configuration of the performance changes (Sup Figs. 1 and 2).

320 Figure 7 illustrates the performance changes as a 3D display by pseudocolor
321 representation as a function of the mean firing frequency and the nonlinearity for three
322 different sampling rates (τ_2 and SNR were fixed at 0.2 s and 5, respectively). HA_time
323 outperformed all of the benchmark algorithms in terms of the F1-score across the entire

324 parameter range (Fig. 7A), except for statistically nondiscriminable MLspike under the
325 high sampling rate and low firing frequency condition (30-60 Hz and ≤ 2 Hz). The three
326 other algorithms, Peeling, MCMC and FRI, exhibited feasible performance only under the
327 high sampling rate, low firing frequency and weak nonlinearity ($\alpha = 0.6-1$) condition.

328



329
330 **Figure 7: Performance of HA_time and benchmark algorithms for simulation data.** A and B,
331 Pseudocolor 3D maps of F1-score and inverse of spike distance as a function of mean firing frequency
332 (abscissa) and nonlinearity (α , ordinate) for the three different sampling rates (10, 30 and 60 Hz). τ_2 and

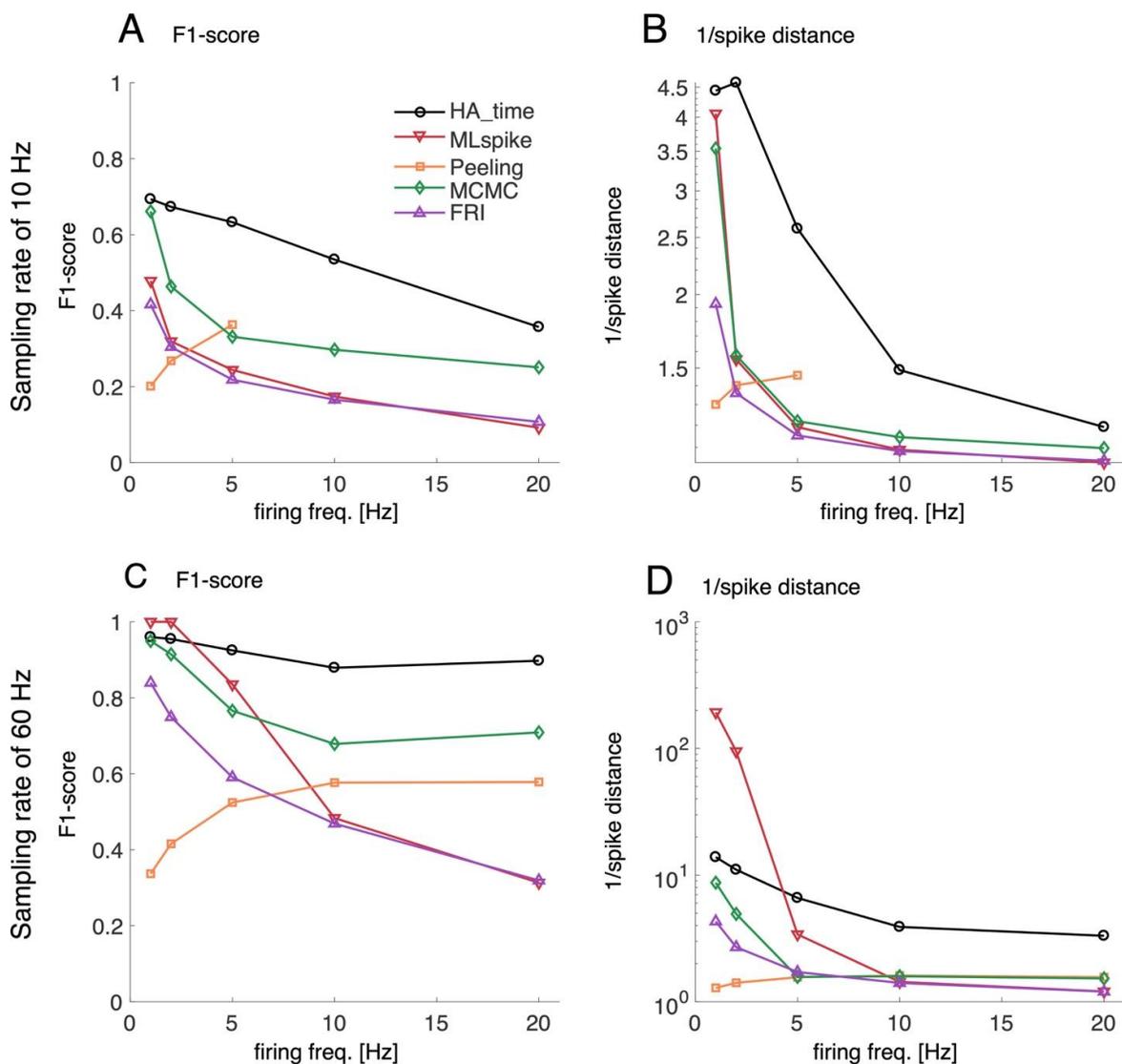
333 SNR were fixed at 0.2 s and 5, respectively. Blank areas indicated that the Peeling algorithm failed to
334 perform in a sufficient time for the cases of high firing frequency and low sampling rate.

335

336 The performance in the temporal precision of spike timings, estimated as the
337 inverse of the spike distance, also showed the same tendency as that for the F1-score.
338 MLspike performed best under the high sampling rate and low firing frequency condition,
339 and the other three benchmark algorithms exhibited feasible performance only under the
340 low firing frequency, weak nonlinearity and high sampling rate condition. Conversely,
341 HA_time outperformed all benchmark algorithms over the entire parameter range except
342 for under the high sampling rate and low firing frequency condition (Fig. 7B).

343 Figure 8 shows profiles of the F1-scores and the inverse of the spike distance
344 values for low (10 Hz) and high sampling rates (60 Hz) as a function of the mean firing
345 frequency (the nonlinearity parameter α was fixed at 1). HA_time outperformed all
346 benchmark algorithms under the low sampling rate condition across the entire range of
347 firing frequencies (1-20 Hz, Figs. 8A and B). The superiority of our algorithm over the
348 benchmark algorithms was also found under the high sampling rate condition across the
349 entire range of spike frequency, except for MLspike, which performed best under the low
350 firing frequency (≤ 2 Hz) condition. However, the performance of MLspike reduced to
351 the level of the other benchmark algorithms as the firing frequency increased (Figs. 8C
352 and D).

353



354

355 **Figure 8: Firing frequency profiles of performance.** A and B, F1-scores and inverse of spike distance
 356 values for HA_time and the benchmark algorithms as a function of firing frequency at a sampling rate of 10
 357 Hz. C and D, those at 60 Hz. The nonlinearity parameter α was fixed at 1. The color conventions for the
 358 algorithms are the same as those in Fig. 6.

359

360 Discussion

361 HA_time aimed to resolve two challenging issues, reliable spike detection and high
 362 spike time precision in the presence of the nonlinearity, slow dynamics and low SNR of
 363 Ca imaging data. The difficulty in achieving this goal arose from the spike dynamics of

364 the target neurons and/or kinematics of the Ca responses that may vary dramatically
365 across brain regions. The slow Ca kinematics and the low SNR may limit the temporal
366 precision of the information conveyed by the Ca imaging data. HA_time overcame this
367 difficulty by combining a generative model for the nonlinearity and Ca dye dynamics with
368 a supervised classifier. It estimated the Ca spike model by Bayesian inference assuming
369 the size and shape constancy of the spike, compensated for the nonlinearity of the Ca
370 responses by nonlinearity analysis, and detected the spikes from the compensated Ca
371 imaging data using the ground truths as supervised information. Hyperacuity precision of
372 spike time estimation was achieved by recalibrating the spike time to minimize the
373 residual errors in the Ca response model prediction using the hyperacuity vernier scale.
374 The combined approach may improve the performance of HA_time in two ways. The Ca
375 response model helped spike detection as well as spike time estimation with higher
376 temporal precision than the sampling resolution, while the supervised learning
377 compensated for fluctuations in the Ca imaging data due to the noise and sampling jitters
378 that are not considered by the generative model.

379 We also developed the generative-model algorithm, hyperacuity Bayes, (see
380 Methods and Supplemental Information) which is a partial algorithm of HA_time. In this
381 algorithm, estimation of the generative model by Bayesian inference with ground-truth
382 information was very robust. However, we ascertained the superiority of the combination
383 approach (i.e. HA_time) in estimating spike timing over the one maximizing the likelihood
384 estimate of the generative algorithm (i.e. hyperacuity Bayes) by the significantly higher
385 F1-score as well as the inverse of the spike distance for the hippocampus and visual
386 cortex data (Fig. S3). Regarding temporal precision, compared with conventional
387 thresholding, HA_time reliably improved up to 20-fold for cases where the sampling rate
388 of Ca imaging was as high as 40-60 Hz. Rapid progress in the development of two-photon
389 imaging techniques with higher sampling rates and faster Ca kinematics may reinforce
390 the superiority of HA_time over conventional methods. We also found that HA_time
391 outperformed the other four hyperacuity algorithms in terms of the F1-score, with feasible
392 statistical significance in the visual cortex and hippocampus data that include a relatively
393 large number of cells. However, for the cerebellum data, in which the number of cells is

394 rather small, HA_time outperformed only the Peeling and MCMC algorithms. HA_time
395 also outperformed all benchmark algorithms in terms of the precision of spike time
396 estimation, which was estimated as the inverse of the spike distance, with statistical
397 significance in visual cortex data. However, HA_time outperformed only MCMC and FRI
398 in the hippocampus data, and outperformed none of the algorithms in the cerebellum
399 data.

400 We conducted a systematic study of algorithm performance on simulation data that
401 covered a broader range of parameters than the experimental conditions, including the
402 mean neuronal firing frequency, the nonlinearity and relaxation time of the Ca dyes and
403 the sampling rate. The performances of the F1-score as well as the inverse of the spike
404 distance values studied as the functions of those parameters pointed to the mean firing
405 frequency, the nonlinearity and the sampling rate as the most important parameters that
406 influenced the changes in performance. In contrast, the SNR and the relaxation time of
407 the dyes are the less important parameters that only influenced the size while not
408 significantly changing the shape of the performance functions. The F1-score and the
409 inverse of spike distance functions highlighted the superiority of HA_time over the other
410 algorithms, showing high scores of spike detection and high spike time precision across
411 the entire range of parameters except for the high sampling rate and low firing rate
412 condition, where MLspike slightly outperformed HA_time. The performance of the other
413 benchmark algorithms remained feasible only under the weak nonlinearity, high sampling
414 rate and low firing frequency condition. Conversely, HA_time maintained high
415 performance under the strong nonlinearity and/or high firing frequency conditions
416 frequently encountered in the cortical cells of behaving animals.

417 The simulation analysis of the performance for HA_time and the benchmark
418 algorithms across a wide range of experimental conditions for two-photon recordings may
419 provide useful information for selection of the best algorithm for given experimental
420 conditions. Although the application of HA_time is limited to cases where ground-truth
421 signals are available, it may also be applicable to cases where simultaneous electrical
422 recordings are unavailable as follows. One may estimate the parameters presently
423 studied for the experimental data of a two-photon recording using a maximum likelihood

424 method, then train HA_time with a newly generated simulation data including spikes for
425 the estimated parameters, and finally use the trained HA_time for spike estimation in the
426 experimental data of the two-photon recording. This approach may benefit from the
427 combination of the generative and supervised approaches, as shown by the present
428 study. In summary, HA_time is useful to improve spike detection and temporal precision
429 in spike time estimation across a wide range of the experimental conditions for two-photon
430 recording in cases where examples of simultaneous ground-truth signals (electrical spike
431 recording) are available.

432

433 **Methods**

434 **Hyperacuity support vector machine (HA_time)**

435 HA_time detects spikes contained in the Ca responses of two-photon recordings in three
436 steps: 1) estimation of the Ca response model by the expectation-maximization (EM)
437 algorithm assuming the shape and size constancy of the spike model, 2) spike detection
438 in the Ca responses by a support vector machine (SVM) assisted by the information of
439 the Ca response matching with the Ca spike model, and 3) spike time estimation for the
440 detected spikes to minimize the errors between the Ca response model prediction and
441 the Ca imaging data using the vernier scale ten-fold finer than the sampling interval of
442 two-photon recordings.

443

444 ***Ca spike model estimation by Bayes inference***

445 We estimated the parameters of the Ca response model assuming that all of the Ca
446 responses in the two-photon recording originate from a unique Ca spike model $g(t, T, \tau)$
447 and vary due to the noise and sampling jitters (SJs)¹⁹.

$$g(t, T, \tau) = \left(1 - \exp\left[-\frac{T-t}{\tau_1}\right]\right) \exp\left[-\frac{T-t}{\tau_2}\right], \quad t \geq T$$

448 $g(t, T, \tau) = 0, \quad t < T$

449 where t , T , $\tau = (\tau_1, \tau_2)$ are the time, the spike onset, and the rise and decay time constants,
450 respectively.

451 We estimated a Ca response model whose parameters are the model amplitude (a),
452 baseline (b_0) and noise (σ) using the EM algorithm, while the time constants of spike
453 response (τ) were estimated by iterative alternate coordinate 1D grid search (see
454 Supplemental Information).

455

456 ***Spike detection by SVM***

457 We conducted spike detection and spike time estimation by an SVM supplemented with
458 the information from the Ca response model to improve the performance of the SVM. We
459 estimated the coincidence scores, determined as the convolution ($dy/dt * dg/dt$) of the
460 first-order derivative (dy/dt) of the Ca signals of a 2-photon recording with that (dg/dt) of
461 the Ca response model (g) estimated by Bayesian inference for the training data, and
462 sampled data segments that exceeded the threshold as the spike candidates. The
463 threshold and the length of data segments (the number of data points before and after
464 the point exceeding the threshold) were optimized according to the F1-score for the
465 training data (cf. cross-validation in Methods). The SVM was trained to classify the spike
466 candidates into spikes or non-spikes by feeding the spike candidates and the coincidence
467 scores as the primary and attribute inputs, respectively, and the electrical spikes (ground
468 truth) as the teaching signals.

469

470 ***Hyperacuity spike time estimation***

471 The trained SVM was used for spike detection in the test data. The pseudo-spike times
472 (PTs) were tentatively determined for the detected spikes as those for which the
473 coincidence score exceeded the threshold. We assumed that the PTs may vary due to
474 the SJ (difference between pseudo- and true spike time) and estimated the SJ to minimize
475 the prediction errors between the Ca response and the Ca response model by
476 systematically changing the SJ according to a vernier scale 10-fold finer than the

477 sampling interval. The true spike time (TT) was calculated as the sum of PT and SJ (cf.
478 the inset of Fig. 1C). We subtracted the trace of the preceding spike from the succeeding
479 one for spike time estimation in cases where the preceding spikes overlap with the
480 succeeding ones.

481

482 **Other benchmark algorithms**

483 We evaluated the performance of four hyperacuity algorithms: the MLspike algorithm²³,
484 the Peeling algorithm¹⁶, the finite-rate innovation method¹⁷ (FRI), and the Monte Carlo
485 Markov chain method²² (MCMC). HA_time and MLspike used the ground-truth signals,
486 given as the electrical spikes for algorithm optimization, whereas the remaining three
487 algorithms did not. We studied how the performance of the three algorithms may be
488 improved in cases where they are supplemented with our parameter settings for the Ca
489 response model and nonlinearity (cf. ^{16,17,22}). To compute the temporal improvement in
490 the hyperacuity algorithms, we also conducted a conventional thresholding algorithm,
491 whose threshold was optimized in the range of 0-4 SD by maximizing the F1-score of the
492 training data.

493

494 **Experimental data sets**

495 We collected simultaneous recordings of electrical and two-photon recording of the Ca
496 signals in three cortical areas (cerebellar, hippocampal and visual cortices) using three
497 different calcium dyes as described below.

498

499 ***Recording of cerebellar Purkinje cell complex spikes***

500 We collected experimental data for the complex spikes of five cerebellar Purkinje cells
501 from the work of ³², where simultaneous two-photon Ca imaging (sampling rate, 7.8 Hz)
502 using multicell bolus loading of Cal-520 dye and extracellular recording (sampling rate,
503 20 KHz) was performed on adult mice.

504

505 ***Recording of hippocampal CA3 neurons***

506 We collected the simultaneous cell-attached recording (sampling rate, 20 KHz) and one-
507 photon images (10 Hz) of Ca responses from nine CA3 pyramidal neurons in organotypic
508 cultured slices of rats stained with OGB-1AM dye³³. The Ca signals were normalized by
509 the peak of the Ca spike model estimated for individual cells.

510

511 ***Recording of the primary visual cortex***

512 We collected the data from the paper of ³⁰. The data set contained simultaneous loose-
513 seal cell-attached patch recordings (sampling rate, 20 KHz) and 2-photon recordings of
514 Ca responses from eleven GCaMP6f-expressing neurons in a behaving mouse visual
515 cortex (sampling rate, 60 Hz).

516

517 **Simulation data**

518 We conducted a simulation of the Ca responses for the three experimental data sets:
519 those of the cerebellar, hippocampal and primary visual cortices. Spike events were
520 generated according to a Poisson distribution whose mean firing rate varied across 1-20
521 Hz. The Ca responses were simulated by convolving the double exponentials with time
522 constants for rise and decay with the spike events. The rise time constant τ_1 was fixed at
523 0.01 s, while the decay time constant τ_2 was varied across 0.2-1 s, corresponding to those
524 for the OGB-1, Cal-520 and GCAMP6f dyes. We introduced the parameter α to reproduce
525 the nonlinearity found in the Ca responses $f(t)$ in the three cortices as

526

527 $f(t) = x(t)^\alpha$, for $x(t) > 1$ and

528 $f(t) = x(t)$; otherwise,

529

530 where $x(t)=g(t)*s(t)$ is the linear response of the Ca spike model given the spike train $s(t)$
531 and the parameter α for saturation ($\alpha<1$) and superlinearity ($\alpha>1$) was varied in the range
532 of 0.2-3, corresponding to the values found in the three experimental data sets. Finally,
533 Gaussian noise was added to reproduce the SNR (3, 5, 10) of the experimental data. For
534 each set of simulation parameters, 500 spike signals in a total of ten cells were generated,
535 and those of five cells were used as the training and test data sets.

536

537 **Performance analysis**

538 For evaluation spike detection performance, the correct hit case was defined as that
539 where the time difference of the estimated spike from the true one was smaller than a
540 window of half the sampling interval, and vice versa for the missing case, and the false-
541 positive case was defined as that where the time difference of the true spike from the
542 estimated one was greater than the time window. For data sets with a high sampling rate
543 (30-60 Hz), the window was relaxed to 50 ms.

544

545 Receiver operating characteristic (ROC) analysis was conducted for these cases as:

- 546 • Sensitivity = Hit / (Hit + misses)
- 547 • Precision = Hit / (Hit + False positive)
- 548 • F1-score = 2 x (Sensitivity x Precision) / (Sensitivity + Precision)

549

550 We estimated the temporal precision of spike time estimation as the inverse of the spike
551 distance, defined as the minimal cost for reconstructing the true spike train from the
552 estimated one, allowing 1 each for deletion or insertion of the spike event and the
553 weighted cost for the shift in the spike time³¹. The spike distance was further normalized
554 by the number of ground-truth spikes. For evaluation of temporal precision improvement,
555 spike time errors were estimated as the absolute time difference between the closest true
556 and reconstructed spikes for all spikes in both spike trains. The hyperacuity improvement

557 was determined as the ratio of the mean spike time errors of conventional thresholding to
558 that of the hyperacuity algorithms.

559

560 **Statistical analysis**

561 Performance analysis of all algorithms on the experimental data was conducted by the
562 one-leave-out cross-validation, where the data of one cell and that of the remaining cells
563 were used for testing and training, respectively.

564

565 All of the performance scores were estimated as the mean \pm 2SEM. To assess statistical
566 significance, we compared the performance of HA_time to that of the benchmark
567 algorithms by a one-sided Wilcoxon signed-rank test and reported the significance level
568 p.

569

570 **Hyperacuity Bayesian Algorithm**

571 We also developed a hyperacuity Bayesian (HB) algorithm by reproducing a basically
572 similar algorithm to that for HA_time. HB is applicable for cases where no ground-truth
573 signals are available, maximizing the likelihood for the Ca signals recorded by two-photon
574 recordings. For cases where the training data with ground-truth signals is provided, HB
575 optimized the model parameters to improve the overall performance of spike time
576 inference (see Supplemental Information).

577

578 **Data availability**

579 The MATLAB® implementation of our algorithm can be found online
580 (https://github.com/hoang-atr/HA_time). The hippocampal and cerebellar cortex data sets
581 used in this work are available from the authors upon reasonable requests.

582

583 **Acknowledgements**

584 This research was conducted under contract with the National Institute of Information and
585 Communications Technology, entitled “Development of network dynamics modeling
586 methods for human brain data simulation systems”. HH and KT were partially supported
587 by the ImPACT Program of Council for Science, Technology and Innovation (Cabinet
588 Office, Government of Japan). HH, YI, and KT were partially supported by JST ERATO
589 (JPMJER1801, "Brain-AI hybrid").

590

591 **Competing interests**

592 The authors declare that no competing interests exist.

593

594 **References**

- 595 1. Kerr, J. N. D. & Denk, W. Imaging in vivo: Watching the brain in action. *Nature*
596 *Reviews Neuroscience* (2008). doi:10.1038/nrn2338
- 597 2. Denk, W., Strickler, J. H. & Webb, W. W. Two-photon laser scanning fluorescence
598 microscopy. *Science* **248**, 73–6 (1990).
- 599 3. Denk, W. *et al.* Anatomical and functional imaging of neurons using 2-photon laser
600 scanning microscopy. *J. Neurosci. Methods* **54**, 151–162 (1994).
- 601 4. Buzsáki, G. Large-scale recording of neuronal ensembles. *Nat. Neurosci.* **7**, 446–
602 51 (2004).
- 603 5. Stosiek, C., Garaschuk, O., Holthoff, K. & Konnerth, A. In vivo two-photon calcium
604 imaging of neuronal networks. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 7319–24 (2003).
- 605 6. Grienberger, C. & Konnerth, A. Imaging Calcium in Neurons. *Neuron* **73**, 862–885
606 (2012).
- 607 7. Hamel, E. J. O., Grewe, B. F., Parker, J. G. & Schnitzer, M. J. Cellular level brain

- 608 imaging in behaving mammals: An engineering approach. *Neuron* **86**, 140–159
609 (2015).
- 610 8. Harris, K. D., Quiroga, R. Q., Freeman, J. & Smith, S. L. Improving data quality in
611 neuronal population recordings. *Nat Neurosci* **19**, 1165–1174 (2016).
- 612 9. Katona, G. *et al.* Fast two-photon in vivo imaging with three-dimensional random-
613 access scanning in large tissue volumes. *Nat. Methods* **9**, 201–208 (2012).
- 614 10. Ji, N., Freeman, J. & Smith, S. L. Technologies for imaging neural activity in large
615 volumes. *Nat. Neurosci.* **19**, 1154–64 (2016).
- 616 11. Tsutsumi, S. *et al.* Structure-function relationships between aldolase C/zebrin II
617 expression and complex spike synchrony in the cerebellum. *J. Neurosci.* **35**, 843–
618 852 (2015).
- 619 12. Yaksi, E. & Friedrich, R. W. Reconstruction of firing rate changes across neuronal
620 populations by temporally deconvolved Ca²⁺ imaging. *Nat. Methods* **3**, 377–383
621 (2006).
- 622 13. Vogelstein, J. T. *et al.* Fast Nonnegative Deconvolution for Spike Train Inference
623 From Population Calcium Imaging Fast Nonnegative Deconvolution for Spike Train
624 Inference From Population Calcium Imaging. *J. Neurophysiol.* **104**, 3691–3704
625 (2012).
- 626 14. Pachitariu, M., Stringer, C. & Harris, K. D. Robustness of Spike Deconvolution for
627 Neuronal Calcium Imaging. *J. Neurosci.* (2018). doi:10.1523/jneurosci.3339-
628 17.2018
- 629 15. Friedrich, J., Zhou, P. & Paninski, L. Fast online deconvolution of calcium imaging
630 data. *PLoS Comput. Biol.* (2017). doi:10.1371/journal.pcbi.1005423
- 631 16. Grewe, B. F., Langer, D., Kasper, H., Kampa, B. M. & Helmchen, F. High-speed in
632 vivo calcium imaging reveals neuronal network activity with near-millisecond
633 precision. *Nat. Methods* **7**, 399–405 (2010).
- 634 17. Oñativia, J., Schultz, S. R. & Dragotti, P. L. A finite rate of innovation algorithm for

- 635 fast and accurate spike detection from two-photon calcium imaging. *J. Neural Eng.*
636 **10**, 046017 (2013).
- 637 18. Greenberg, D. S., Houweling, A. R. & Kerr, J. N. D. Population imaging of ongoing
638 neuronal activity in the visual cortex of awake rats. *Nat. Neurosci.* **11**, 749–751
639 (2008).
- 640 19. Lütcke, H., Gerhard, F., Zenke, F., Gerstner, W. & Helmchen, F. Inference of
641 neuronal network spike dynamics and topology from calcium imaging data. *Front.*
642 *Neural Circuits* **7**, 201 (2013).
- 643 20. Sebastian, J. *et al.* GDspike: an accurate spike estimation algorithm from noisy
644 calcium fluorescence signals. in *IEEE International Conference on Acoustics,*
645 *Speech and Signal Processing* 1043–1047 (2017).
646 doi:10.1109/ICASSP.2017.7952315
- 647 21. Vogelstein, J. T. *et al.* Spike inference from calcium imaging using sequential Monte
648 Carlo methods. *Biophys. J.* **97**, 636–655 (2009).
- 649 22. Pnevmatikakis, E. A. *et al.* Simultaneous Denoising, Deconvolution, and Demixing
650 of Calcium Imaging Data. *Neuron* **89**, 299 (2016).
- 651 23. Deneux, T. *et al.* Accurate spike estimation from noisy calcium signals for ultrafast
652 three-dimensional imaging of large neuronal populations in vivo. *Nat. Commun.* **7**,
653 12190 (2016).
- 654 24. Theis, L. *et al.* Benchmarking Spike Rate Inference in Population Calcium Imaging.
655 *Neuron* **90**, 471–482 (2016).
- 656 25. Sasaki, T., Takahashi, N., Matsuki, N. & Ikegaya, Y. Fast and accurate detection of
657 action potentials from somatic calcium fluctuations. *J. Neurophysiol.* **100**, 1668–76
658 (2008).
- 659 26. Berens, P. *et al.* Community-based benchmarking improves spike rate inference
660 from two-photon calcium imaging data. *PLoS Comput. Biol.* (2018).
661 doi:10.1371/journal.pcbi.1006157

- 662 27. Stringer, C. & Pachitariu, M. Computational processing of neural recordings from
663 calcium imaging data. *Curr. Opin. Neurobiol.* **55**, 22–31 (2019).
- 664 28. Peron, S., Chen, T. W. & Svoboda, K. Comprehensive imaging of cortical networks.
665 *Current Opinion in Neurobiology* (2015). doi:10.1016/j.conb.2015.03.016
- 666 29. Wilt, B. A., Fitzgerald, J. E. & Schnitzer, M. J. Photon shot noise limits on optical
667 detection of neuronal spikes and estimation of spike timing. *Biophys. J.* (2013).
668 doi:10.1016/j.bpj.2012.07.058
- 669 30. Chen, T.-W. *et al.* Ultrasensitive fluorescent proteins for imaging neuronal activity.
670 *Nature* (2013). doi:10.1038/nature12354
- 671 31. Victor, J. D. & Purpura, K. P. Metric-space analysis of spike trains: theory,
672 algorithms and application. *Netw. Comput. Neural Syst.* **8**, 127–164 (1997).
- 673 32. Tada, M., Takeuchi, A., Hashizume, M., Kitamura, K. & Kano, M. A highly sensitive
674 fluorescent indicator dye for calcium imaging of neural activity in vitro and in vivo.
675 *Eur. J. Neurosci.* (2014). doi:10.1111/ejn.12476
- 676 33. Takahashi, N., Sasaki, T., Usami, A., Matsuki, N. & Ikegaya, Y. Watching neuronal
677 circuit dynamics through functional multineuron calcium imaging (fMCI). *Neurosci.*
678 *Res.* **58**, 219–225 (2007).
- 679
- 680

681 **Supplemental Information**

682

683 **Hyperacuity Bayesian algorithm**

684 We also developed a hyperacuity Bayesian (HB) algorithm for spike detection and
685 spike time estimation, maximizing the estimate likelihood for the cases where ground-
686 truth signals are not available. The HB algorithm included supervised and unsupervised
687 versions. The supervised version reproduces essentially similar procedures, such as
688 estimation of the Ca response model, spike detection and spike time estimation using the
689 model information, to those for HA_time by the Bayesian algorithm. Probabilistic models
690 and the EM algorithm are described in the first two sections, and detailed procedures of
691 the two versions of the HB algorithm are described in the next two sections.

692

693 ***Data structure and probabilistic model***

694 Let us suppose that K data segments were sampled from data by thresholding while
695 leaving the rest of the data (y_{rest}).

$$Y = [\{y_k | k = 1 : K\}; \{y_{rest}(t) | t = 1 : N_{rest}\}]$$

$$y_k = \{y_k(n) | n = 1 : N\}$$

696

697

698 where $y_k(n)$ is the sampled data at the sampling time $t_{k,n} = (n-1)dt_0 + t_{k,1}$ of the k-th window
699 and dt_0 is the sampling step of the observed data with the sampling frequency $f_0 = 1/dt_0$.

700

701 A probabilistic model for spike states is given by

$$\begin{aligned} \log(P(Y|A, S, T, \phi)) &= \sum_{k=1}^K \log(P(y_k|A_k, S_k, T_k, \phi)) + \sum_{t=1}^{N_{rest}} \log(P(y_{rest}(t)|\phi)) \\ \log(P(y_k|A_k, S_k, T_k, \phi)) &= -\frac{N}{2\sigma} E(y_k, A_k, T_k, \phi) - \frac{N}{2} \log(\sigma) + const \\ E(y_k, A_k, S_k, T_k, \phi) &= \frac{1}{N} \sum_{n=1}^N \left(y_k(n) - A_k \cdot \sum_{N_k=1}^{N_{max}} S_k \cdot \left[\sum_{s=1}^{N_k} a \cdot g(t_{k,n}, T_k(s), \tau) + b \right] \right)^2 \\ \log(P(y_{rest}(t)|\phi)) &= -\frac{1}{2\sigma} (y_{rest}(t) - b_0)^2 - \frac{1}{2} \log(\sigma) + const \end{aligned}$$

702

703 , where A_k is a spike indicator variable, which represents the presence or absence of
 704 spikes in the k -th window, and takes a binary value (0 or 1). S_k represents a spike state
 705 in the k -th window and takes a binary vector value (Potts spin variable)

$$\sum_{n=1}^{N_{max}} S_k(n) = 1$$

706

$$S_k(n) = 0 \text{ or } 1$$

707

708 $S_k(N_k) = 1$ means that there are N_k spikes in the k -th window. The maximum number of
 709 spikes in a window is assumed to be N_{max} . T_k is a set of spike times in the k -th window,
 710 and $T_k(s)$ is the onset time of the s -th spike. a represents the amplitude of the spike
 711 response function. b and b_0 represent the bias in spike and no spike regions, respectively.
 712 σ is the variance of the Gaussian noise. A set of global parameters that is assumed to be
 713 common for all spikes and the rest of the data is denoted by $\phi = (a, \tau, b, b_0, \sigma)$.

714

715 We assume the hierarchical noninformative priors as

$$\begin{aligned} P_0(A_k) &= 1/2 \\ P_0(S_k(N_k) = 1|A_k = 1) &= 1/N_{max} \\ P_0(T_k|S_k(N_k) = 1, A_k = 1) &= 1/N_{conf}(N_k) \\ P_0(\phi) &= const \end{aligned}$$

716

717 , where $N_{conf}(N_k)$ represents the number of configurations of T_k for the N_k spike case.

718

719 **Expectation maximization (EM) algorithm**

720 In the E-step, the posterior probability of the spike state for the current estimate of the
721 model parameters $\phi = (a, \tau, b, b_0, \sigma)$ for combination of the log-likelihood, the noninformative
722 prior and the joint probability for a spike state in the k-th window is given by:

723
$$P(y_k, A_k, S_k, T_k | \phi) = P(y_k | A_k, S_k, T_k, \phi) P_0(T_k | S_k, A_k) P_0(S_k | A_k) P_0(A_k)$$

724 and marginal probability is given by:

725
$$P(y_k | \phi) = \sum_{A_k} \sum_{S_k} \sum_{T_k} P(y_k, A_k, S_k, T_k | \phi)$$

726 The posterior probability for a spike state is then calculated as

727
$$P(A_k, S_k, T_k | y_k, \phi) = \frac{P(y_k, A_k, S_k, T_k | \phi)}{P(y_k | \phi)}$$

728

729 In the M-step, the model parameters were updated to a new value ϕ_{new} by maximizing
730 the Q-function, defined as

731
$$Q(\phi_{new}) = \sum_k \sum_{A_k} \sum_{S_k} \sum_{T_k} (\log(P(A_k, S_k, T_k, y_k | \phi_{new}))) P(A_k, S_k, T_k | y_k, \phi) \\ + \sum_t (\log(P(y_{rest}(t) | \phi_{new})))$$

732 by solving the maximum condition

733
$$\partial Q(\phi_{new}) / \partial \phi_{new} = 0$$

734

735 **Spike estimation for the supervised version of HB**

736 We first extracted continuous regions in which the signals exceed the threshold.
737 Here, the threshold was optimized to maximize the true-positive cases and minimize the
738 false-positive cases referring to the ground truth given by the electrical spikes. Each
739 region is then segmented into fixed-length data segments of 8 points. We divided two
740 overlapping data segments, whose onset intervals were less than the length of the data
741 segment (8 points), into three divided nonoverlapping segments. For example, two
742 segments of points #1-10 and #5-14 were divided into three nonoverlapping segments of
743 points #1-4, #5-10 and #11-14.

744 Next, we utilized the training data to estimate the model parameters and the
745 classifier for posterior probability. The spike model parameters including the spike
746 amplitude (a), biases (b , b_0) and noise variance σ were estimated by the EM algorithm
747 (see above), while time constants of the spike response (τ) were estimated by iterative
748 alternate coordinate 1D grid search because the log-likelihood with respect to τ is highly
749 nonlinear. The maximum number of spikes contained in single data segments was
750 estimated for the 95-percentile value of the spike number histogram of the training data
751 segments. The posterior probability of the spike state for training data was estimated for
752 each data segment assuming that they are independent of each other. For overlapping
753 segments, posterior probabilities were integrated among the overlapping segments by
754 Bayes inference. We used a multinomial classifier to predict the number of spikes from
755 the posterior probability.

756 Finally, we estimated the spike number and spike onset time for test data. The
757 data segments were sampled for the test data by thresholding whose threshold was
758 optimized for the training data, and the posterior probability of the spike state for each
759 segment was calculated in the same way as for the training data. The number of spikes
760 was estimated from the posterior probability for the number of spikes using the
761 multinomial classifier trained for the training data, and spike onset times were estimated
762 with the hyperacuity time step by maximizing the log-likelihood for the estimated number
763 of spikes. The contributions of preceding spikes were subtracted from the data signal, as
764 was done for HA_time.

765

766 ***Spike estimation by the unsupervised version of HB***

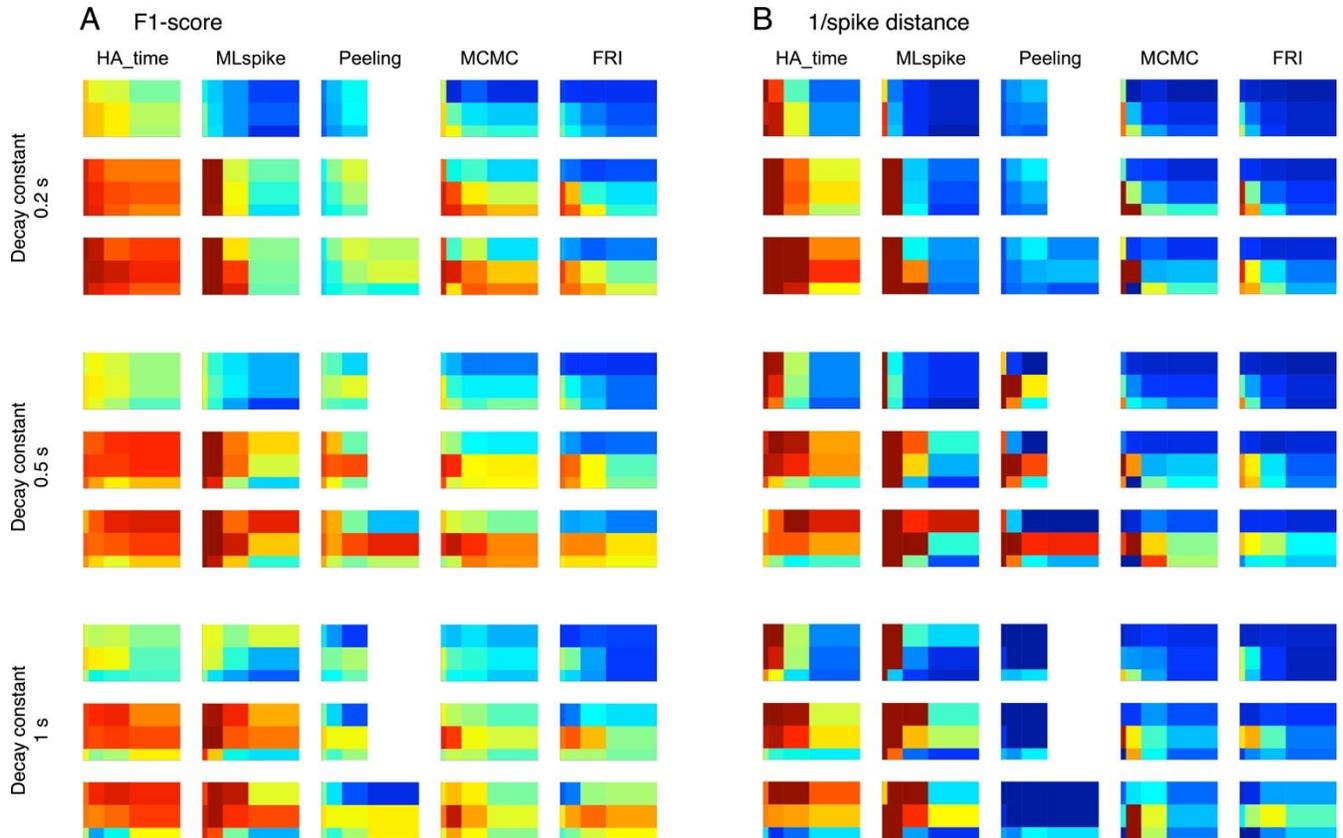
767 The unsupervised version of HB conducted the spike estimation in essentially the
768 same way as that for the supervised version, except for the algorithm used to estimate
769 the spike state. We conducted Bayesian inference assuming that the initial spike state for
770 each data segment contains only one spike whose waveform represents the spike
771 response. We optimized the threshold and $\tau = (\tau_1, \tau_2)$ that maximize the log-likelihood and

772 estimated the number of spikes for each data segment that gave the maximum posterior
773 probability for the given number of spikes.

774

775 Supplemental Figures

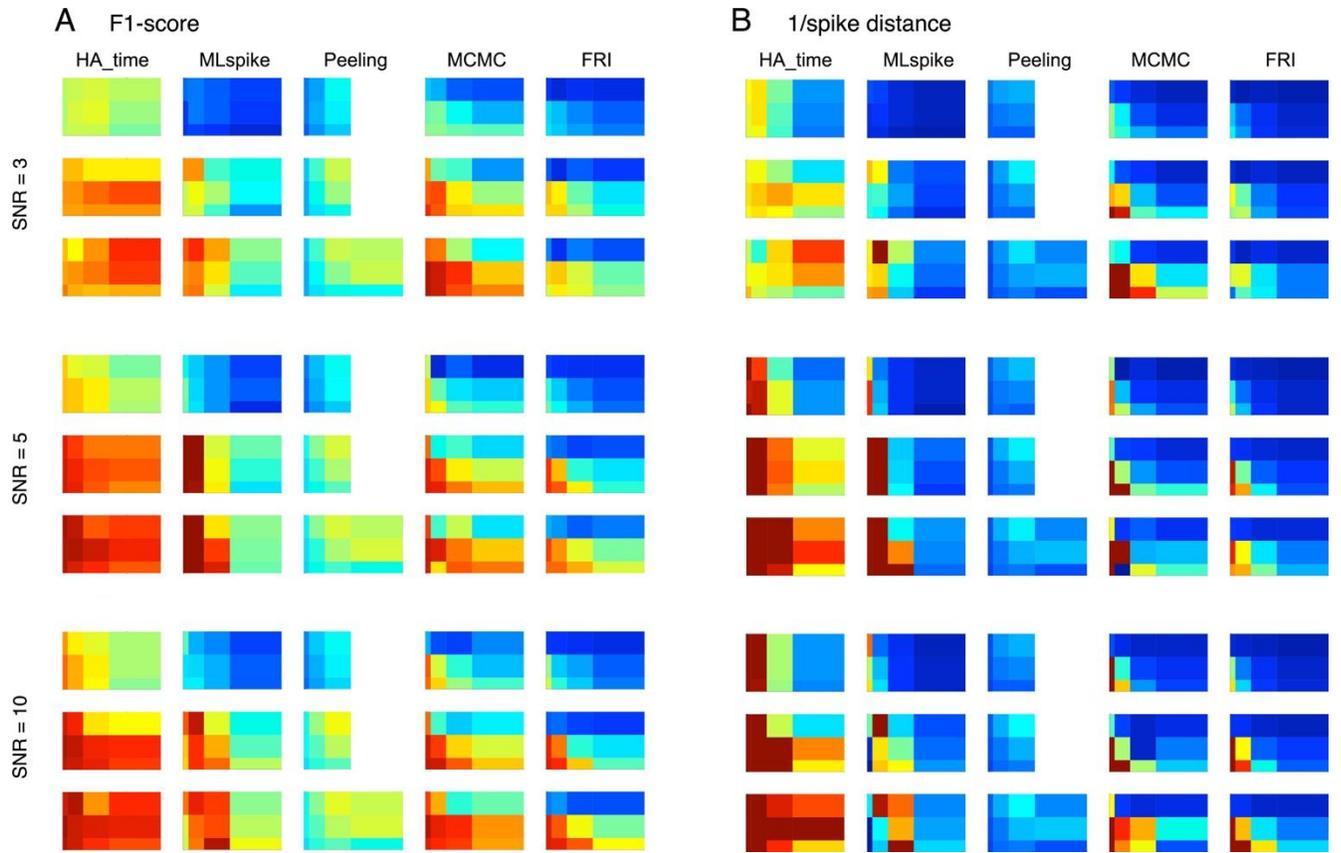
776



777

778 **Supplementary Figure 1: Performance of HA_time and benchmark algorithms on simulation data**
779 **for variation in the decay time constant of the Ca response model.** A and B, performance of F1-score
780 and inverse of spike distance for three different decay time constants compared to that in Fig. 5 for the Ca
781 response model (0.2, 0.5 and 1 s) and three different sampling rates (10, 30 and 60 Hz). The ordinate,
782 abscissa and calibration scale of the pseudocolor maps are the same as those in Fig. 5.

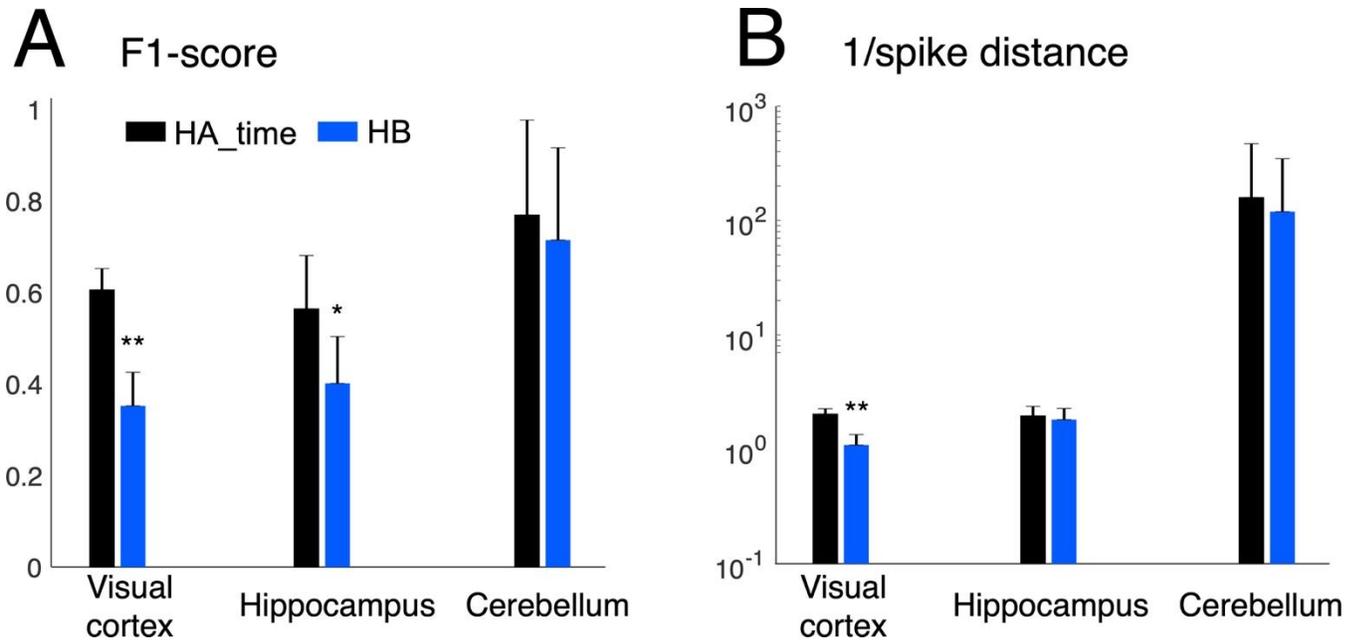
783



784

785 **Supplementary Figure 2: Performance of HA_time and benchmark algorithms on simulation data**
786 **for variation in the SNR of Ca response signals. A and B, performance of F1-score and inverse of spike**
787 **distance similar to Fig. S1 but for three different SNRs of the Ca response signals.**

788



789

790 **Supplementary Figure 3: Performance benchmark of HA_time and HB for experimental data.** A and
791 B, F1-score and inverse of the spike distance for HA_time (black columns) and HB (blue). Conventions are
792 the same as those in Fig. 6.