MOD: A novel machine-learning optimal-filtering method for accurate and efficient detection of subthreshold synaptic events *in vivo*

Xiaomin Zhang[1,2]*, Alois Schlögl[1]*, David Vandael[1,3] & Peter Jonas[1**]

[1]IST Austria (Institute of Science and Technology Austria), Am Campus 1, A-3400 Klosterneuburg, Austria

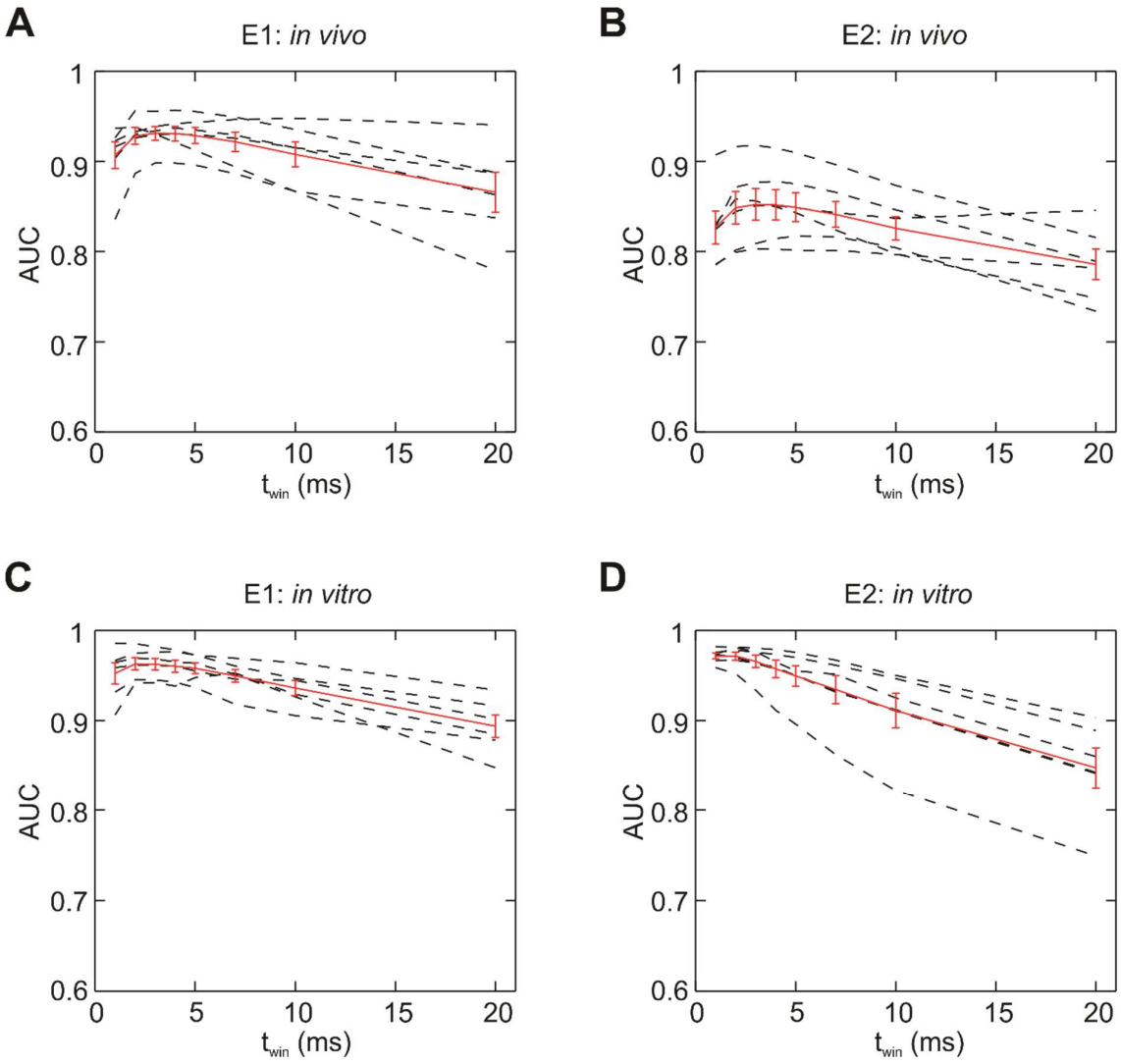
[2]Present address: Brain Research Institute, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

[3]Present address: Netherlands Institute for Neuroscience, Royal Netherlands Academy for Arts and Sciences (KNAW), Meibergdreef 47, 1105 BA, Amsterdam, The Netherlands

* Equal contribution

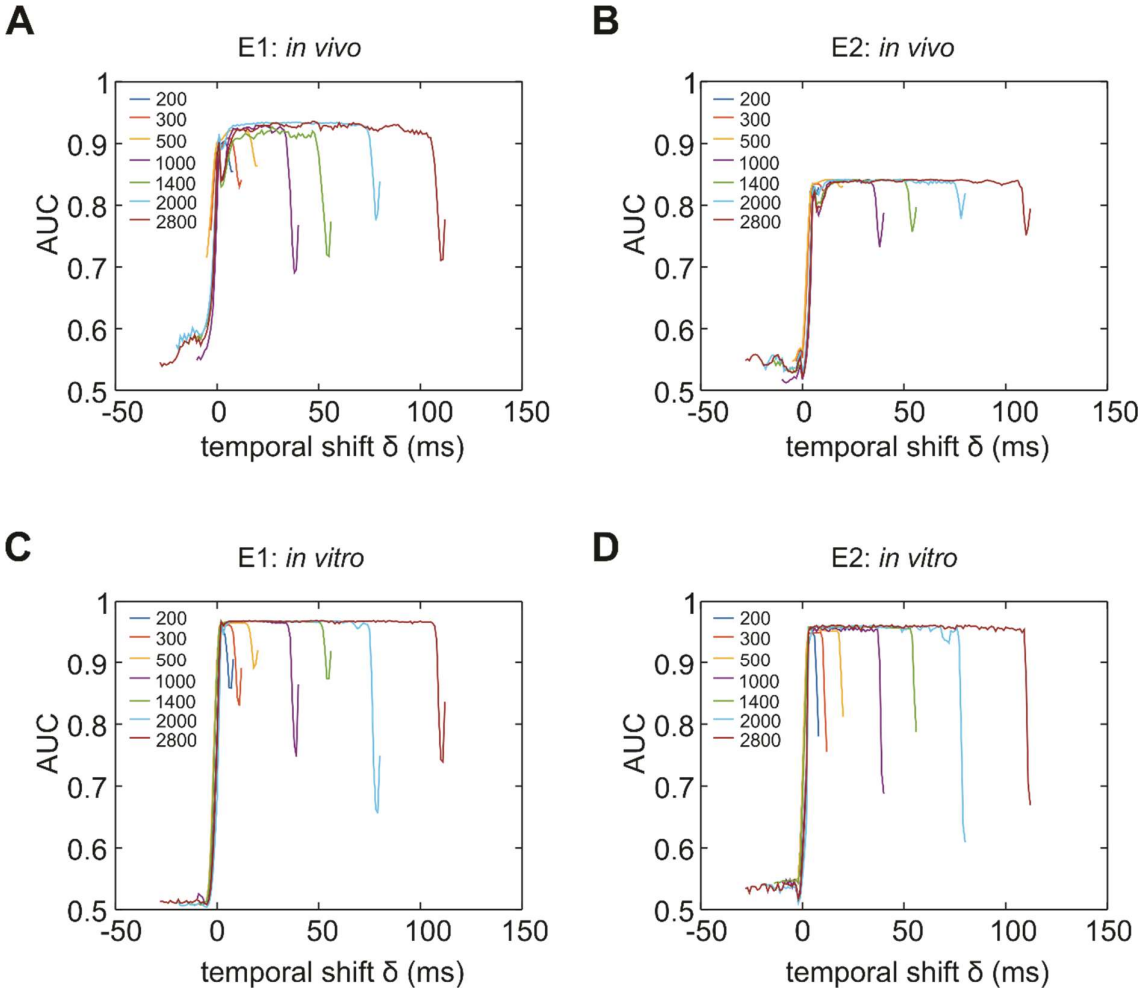
****Corresponding author:**
Dr. Peter Jonas
IST Austria
Am Campus 1
A-3400 Klosterneuburg, Austria
Phone: +43-2243-9000-3700
E-mail: peter.jonas@ist.ac.at

Fig. S1. The optimal time window for manual scoring (t_{win}).



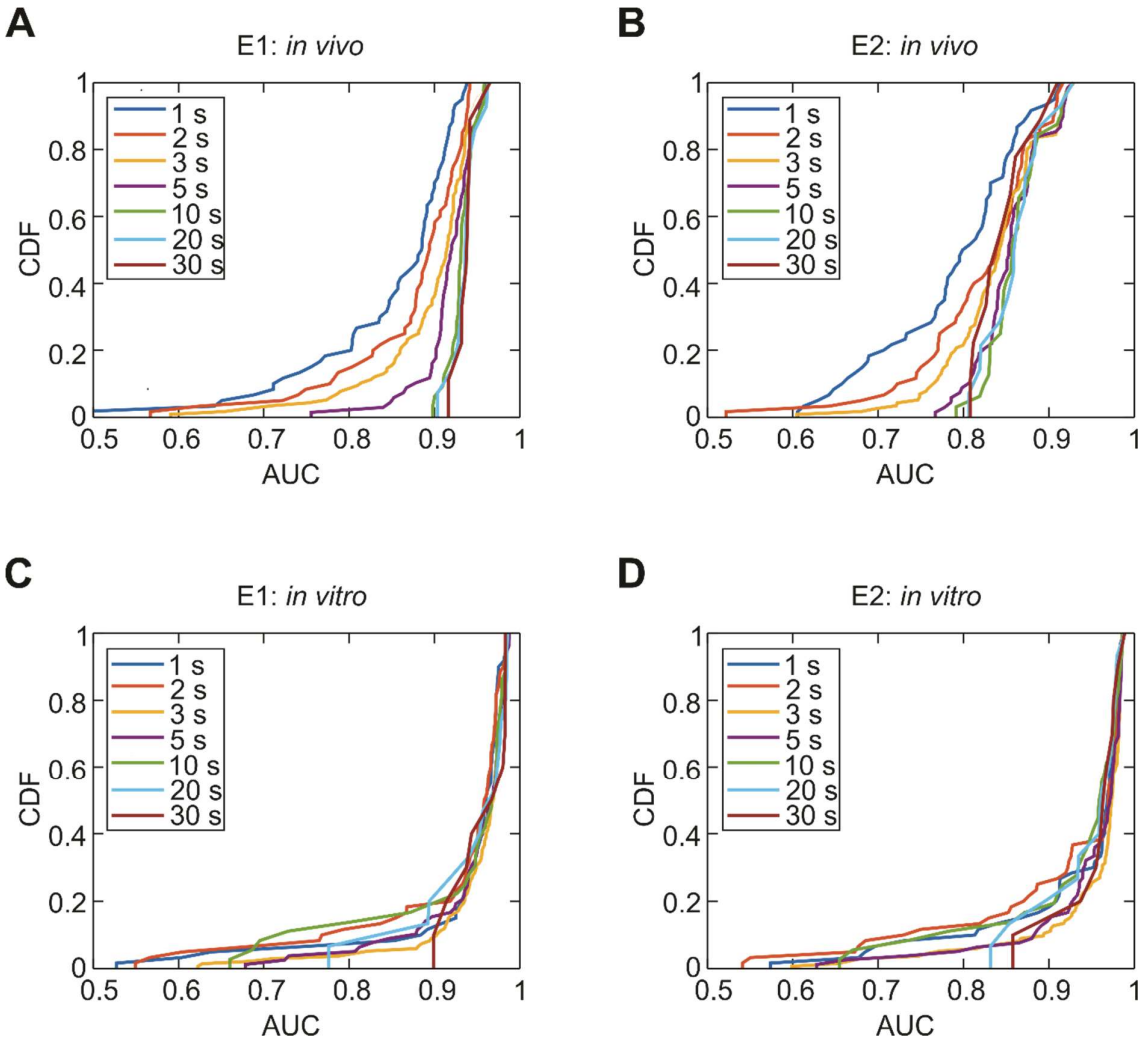
Analysis of the dependence of AUC on the duration of the manual scoring expansion window (t_{win}). Data were obtained from *in vivo* recordings (**A, B**) or *in vitro* recordings (**C, D**) and scored by either expert E1 (**A, C**) or expert E2 (**B, D**). Note that the optimum t_{win} value was 3–4 ms under all conditions. Red curve, mean ± SEM; dashed lines, data from individual experiments. The optimum value of t_{win} defines the temporal resolution of the entire method, including manual scoring and automatic detection. Results shown were computed using the cross-validation scheme "A1B2–A2B1".

Fig. S2. The optimal time shift δ and influence of filter length of the Wiener filter.



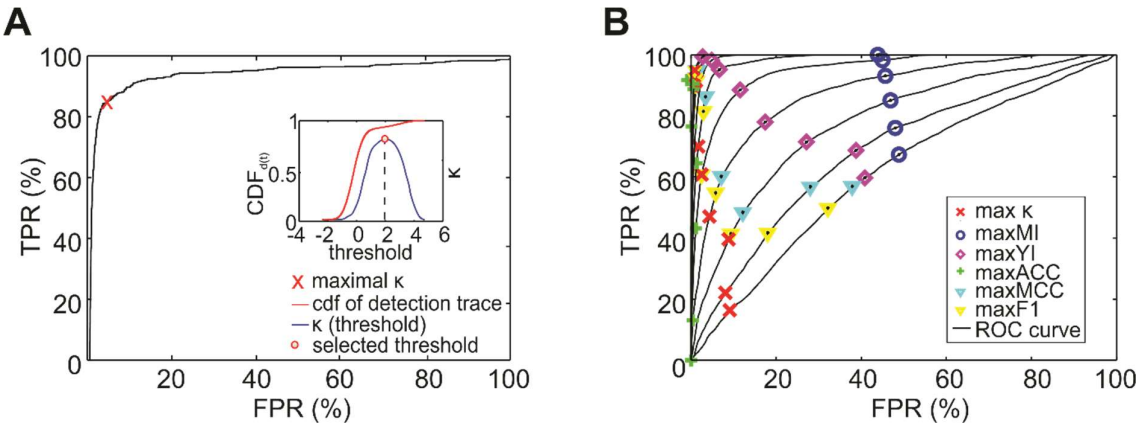
Analysis of the dependence of AUC on the time shift δ for various values of the filter length (200, 300, 500, 1000, 1400, 2000, and 2800 samples, sampling frequency 25000 Hz). δ = 0 corresponds to the time points of manual scoring. Data were obtained from *in vivo* recordings (**A, B**) or *in vitro* recordings (**C, D**) and scored by either expert E1 (**A, C**) or expert E2 (**B, D**). Overall, the filter length has only minor effects on the AUC. Note that the optimum time shift was ~20 ms for the *in vivo* data sets and ~5 ms for the *in vitro* data sets. This is likely to be caused by the slower time course of EPSPs in the *in vivo* data in comparison to the faster time course of the EPSCs in the *in vitro* data. Also note that delay values were slightly different between expert scorings (E1 versus E2), because experts marked synaptic events at slightly different time points. The results shown were computed using the general classifier.

Fig. S3. The minimal length of scored traces required for reliable detection.



Cumulative distribution of AUC values for different duration of the scoring period used for training of MOD. Data were obtained from *in vivo* recordings (**A, B**) or *in vitro* recordings (**C, D**) and scored by either expert E1 (**A, C**) or expert E2 (**B, D**). Note that AUC values were consistently above 0.8 for scoring periods of ≥ 10 s. In contrast, a fraction of analysis data points showed AUC values of < 0.8 for shorter scoring periods of 1–5 s. Thus, MOD analysis based on shorter scoring periods is possible, but occasionally unreliable.

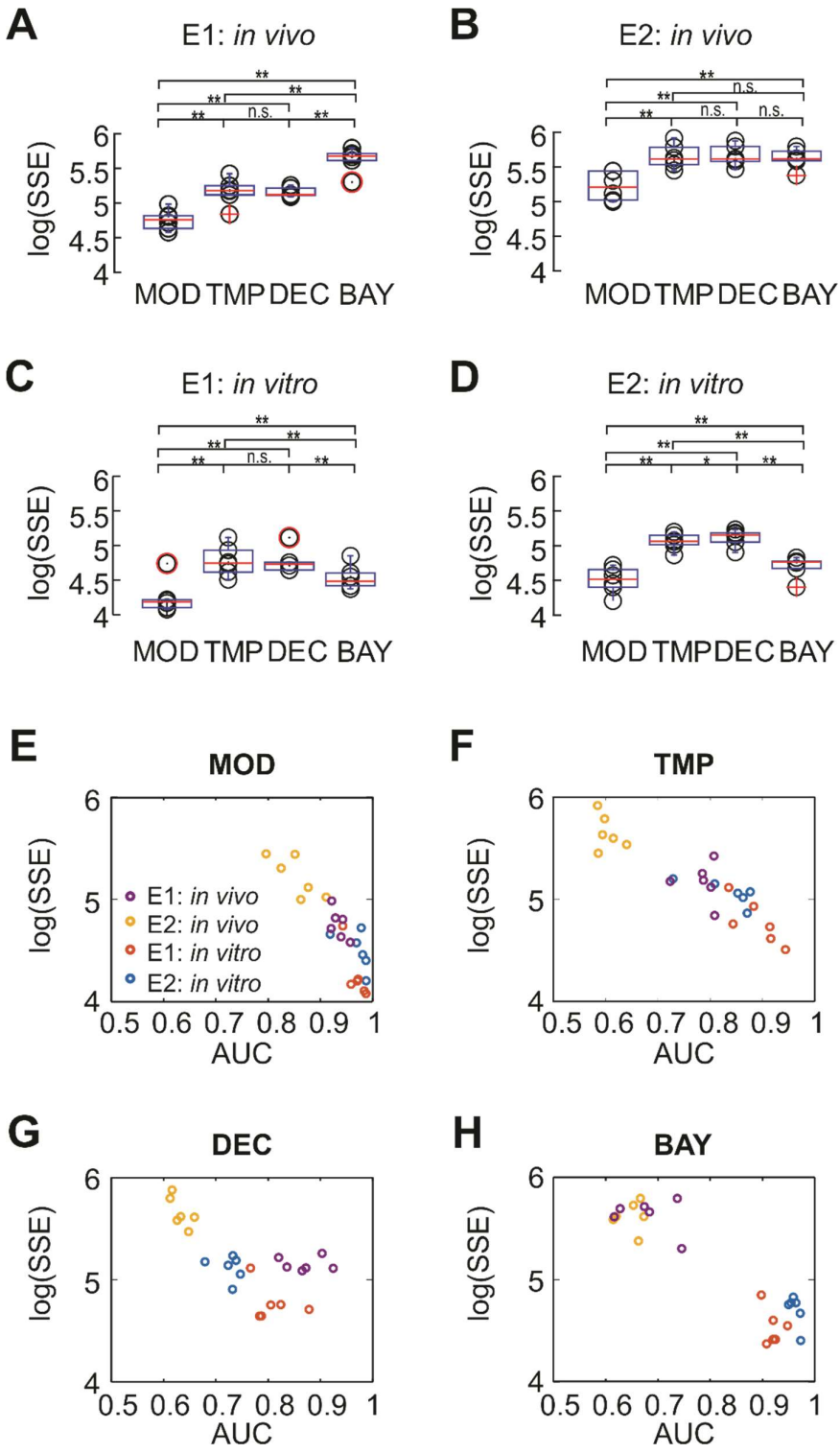
Fig. S4. Comparison of different criteria for determining detection threshold.



(**A**) Plot of TPR against FPR for a synthetic data set (ROC curve). The signal-to-noise ratio was 3, i.e. 9.5 dB in this simulation. The AUC was 0.970. The "x" in the ROC curve indicates the point on the ROC curve that corresponds to maximal κ, implying the optimal detection threshold. Inset shows the dependence of Cohen's κ on the threshold, peak value of κ, and cumulative distribution of data points in the detection trace.

(**B**) ROC curves for simulated data. The signal-to-noise ratio SNR = $20*\log_{10}$(peak amplitude/noise rms) was varied from −52 dB to 26 dB in steps of 7.8 dB. Note that larger signal-to-noise ratios yield ROC curves with a larger AUC. The maxima of different performance metrics (Cohen's κ, mutual information (MI), Youden index (YI), accuracy (ACC), Matthews correlation coefficients (MCC), and F1) are depicted.

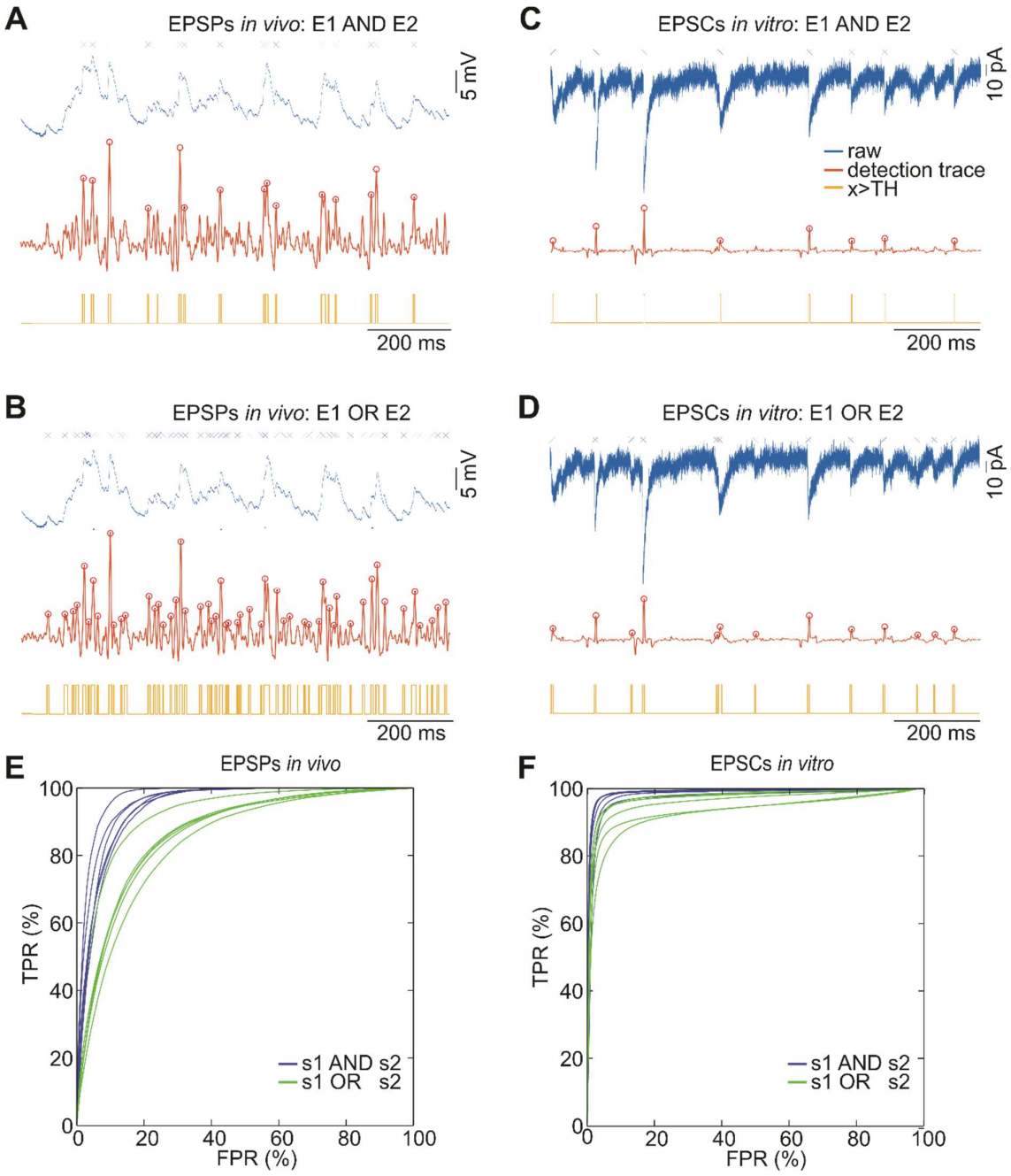
Fig. S5. Comparison of detection methods using an SSE-based goodness-of-fit metric and relationship between SSE and AUC.



(**A, B**) Comparison of the performance of MOD against previously published methods (TMP, template-fit; DEC, deconvolution; BAY, Bayesian detection) for 6 *in vivo* data sets. Note that SSE is smallest for MOD, indicating MOD outperforms the traditional methods TMP, DEC, and BAY for EPSPs recorded *in vivo*.

(**C, D**) Comparison of the performance of MOD against TMP; DEC; BAY for 6 *in vitro* data sets. Note that MOD outperforms the traditional approaches TMP and DEC for EPSCs recorded *in vitro* (* indicates $P < 0.05$). Results shown were computed using the cross-validation scheme "A1B2–A2B1" with $t_{win} = 4$ ms.
(**E**) Scatter plot of $\log_{10}$(SSE) versus AUC for *in vivo* and *in vitro* data sets scored by experts E1 and E2 for MOD.
(**F–H**) Scatter plot of $\log_{10}$(SSE) versus AUC for TMP (F), DEC (G), and BAY (H).
SSE and AUC were highly correlated in all data sets, implying that both may allow judgement of the goodness of fit. Pearson's correlation coefficients were 0.6343 (E), 0.8461 (F), 0.7213 (G), and 0.9608 (H), and corresponding P values were 8.7279×10^{-4} (E), 1.9027×10^{-7} (F), 6.9701×10^{-5} (G), and 9.5923×10^{-14} (H). Note, however, that SSE is dependent of the choice of an optimal threshold, whereas AUC is independent on any assumptions about threshold.

Fig. S6. Using MOD with combined expert scoring.



(**A–D**) One-second segments of scored *in vivo* data (A, B) and *in vitro* data (C, D) for AND combination of scorings (A, C) or OR combination of scorings (B, D). Top, raw data, together with the scoring from one expert. Center, raw detection trace from the MOD method; the markers (○) indicate peaks above threshold. Bottom, binary detection trace obtained by applying the optimum threshold to the raw detection trace.

(**E, F**) ROC curve analysis of detection performance for *in vivo* data (E) and *in vitro* data (F). Blue lines, AND combination; green lines, OR combination.

Table S1. Properties of manually scored data sets.

Parameter	*in vivo* data set	*in vitro* data set	Additional information
Total scoring duration	71.0 ± 10.1 s	73.7 ± 8.4 s	
Total number of scored events	2100 ± 683	811 ± 305	
Average event frequency	30.0 ± 9.3 Hz	11.4 ± 5.6 Hz	
Number of scored cells	6	6	
Number of scorings	12	12	Expert E1 with *in vivo* background, expert E2 with *in vitro* background

In vivo data set was obtained by making whole-cell current-clamp recordings of EPSPs in dentate gyrus GCs near the resting potential in head-fixed mice running on a linear belt (for details, see Methods).

In vitro data set was obtained using presynaptic cell-attached recording from hippocampal mossy fiber terminals and simultaneous postsynaptic whole-cell voltage-clamp recordings of EPSCs from CA3 pyramidal neurons in acute hippocampal slices (for details, see Methods).

Table S2. Computation time for individuals steps of MOD.

	Computation time (mean ± SD)	Reference	Conditions
Computing auto-correlation functions	14.14 ± 0.32 s	Eq. 6a	normalized to 30 s of training data
Computing cross-correlation functions	35.33 ± 0.79 s	Eq. 6b	normalized to 30 s of training data
Solving Wiener-Hopf equation	0.028 ± 0.001 s	Eq. 5	
Raw data filtering	14.59 ± 0.26 s	Eqs. 2 and 8	normalized to 600 s of raw data
Computing ROC, AUC, and κ	0.073 ± 0.044 s	Fig. 2C, F; Fig. 3C; Fig. 4C	normalized to 600 s of raw data

The results were computed on a Supermicro computer with an X9DRT mainboard, and Intel Xeon CPU (E5-2670 v2 @ 2.60GHz, Sandy Bridge, with hyper-threading enabled). The machine was running under the operating system "Debian(9.6)/Stretch", and the computations were done with Octave 4.4.0 on a single CPU core.

Computation time is wall clock time in seconds. Twelve data sets were analyzed, and the cross-validation scheme (A1B2–A2B1) was used, thus 24 results were computed. The computational time of each data set was normalized to a length of 30 s of training data, and 600 s of total raw data.