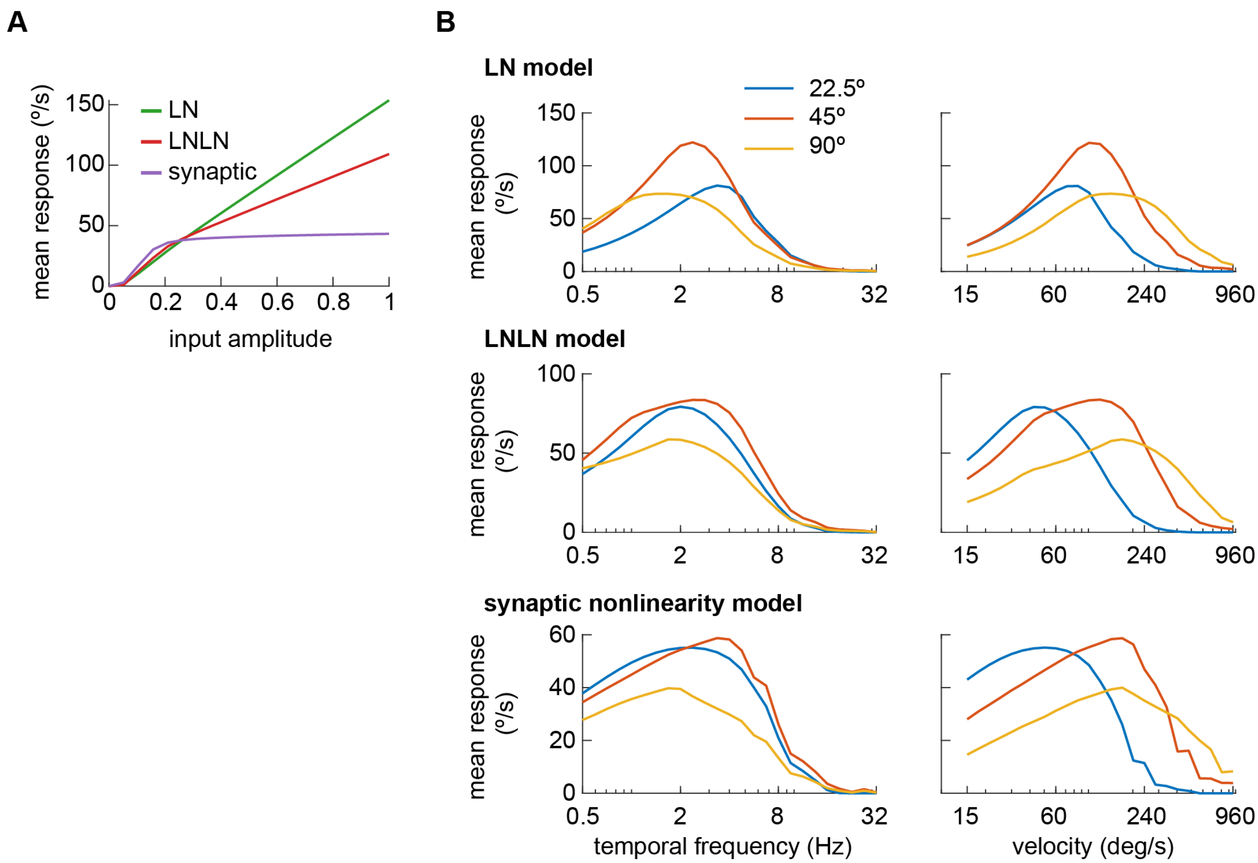

Figure S1. Tuning of trained models to sinusoids of different amplitudes and different spatiotemporal frequencies, related to Figure 3.

A) Mean responses of trained LN, LNLN, and synaptic nonlinearity models to sinusoidal inputs with different amplitudes. The models are those shown in Figure 3. Sinusoids had wavelength 45° and temporal frequency 1 Hz.

B) Mean response of trained LN, LNLN, and synaptic nonlinearity models to sinusoidal inputs with different spatial and temporal frequencies, as noted in figure. The models are those shown in Figure 3. The LNLN and synaptic nonlinearity models are quite temporal frequency-tuned, while the LN model tuning to temporal frequency is more modest. The training task was to predict velocity with minimum squared error, so it may not be surprising that some models, like the trained LN model, appear more velocity tuned.

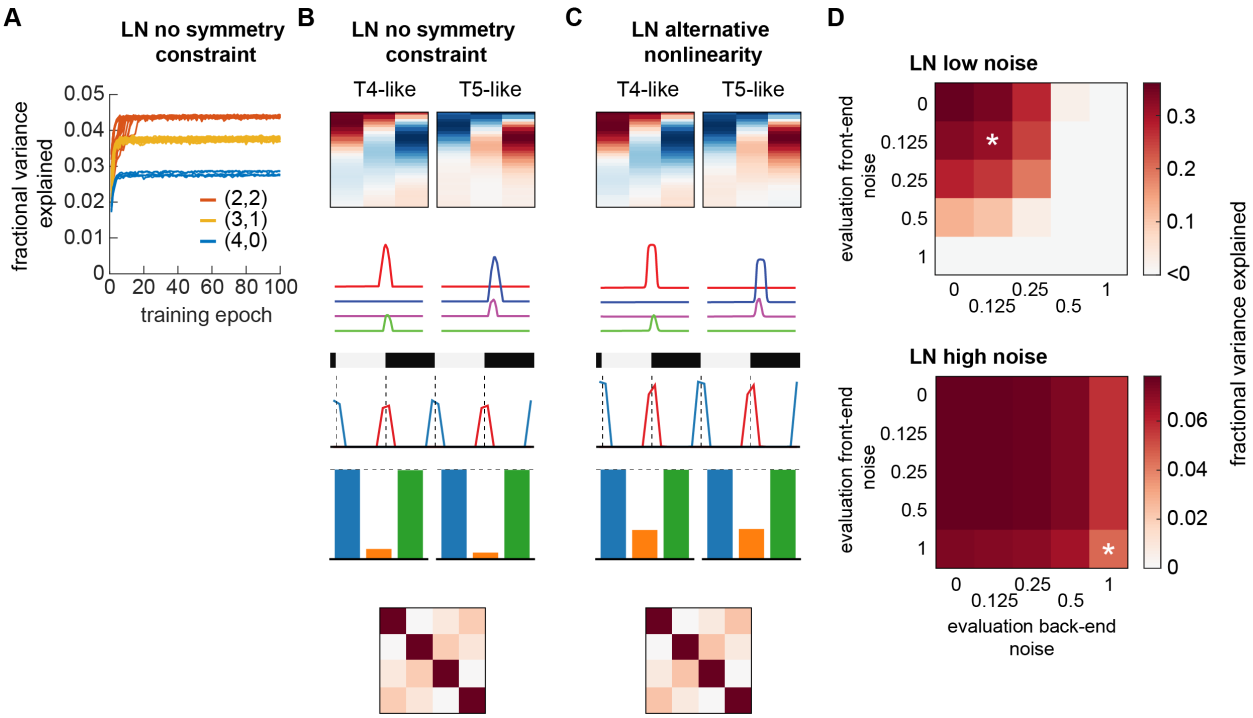

Figure S2. Trained solutions with different constraints and general effects of noise on solutions, related to Figure 4.

A) Variance accounted for by different LN model solutions when mirror symmetry between the unit pairs was not enforced. Four units were trained using 3 independent filters each, for 12 total filters, relaxing the requirement for mirror-symmetry and subtraction found in regular training. Solutions are colored by the number of units tuned to each direction. For instance (3,1) indicates 3 units tuned to one direction, 1 to the opposite direction. The best performing models had two units tuned to each direction (i.e., (2,2)). Models were trained in high noise regime, as in Figure 3.

B) LN model characterization of a top-performing solution when symmetry requirements were relaxed (as in (A)). The best performing solutions still obtain mirror symmetry in their units and still split into ON- and OFF-edge detectors, with response properties similar to models trained with symmetry enforced. Model trained in high noise regime, as in Figure 3. Compare with results in Figures 3 and 4.

C) LN model characterization when trained with a saturating nonlinearity, $\phi(x) = \frac{1}{2}(1 + \tanh x)$, instead of the rectified linear unit used elsewhere. Model response properties remain very similar. Model trained in high noise regime, as in Figure 3. Compare with results in Figures 3 and 4.

D) Variance accounted for by LN models trained in low noise ($\eta = \sigma = 1/8$, *top*) and high noise ($\eta = \sigma = 1$, *bottom*), evaluated using holdout data under different noise conditions. The high-noise trained LN model has similar performance over many noise levels, while the low-noise trained LN model performance degrades sharply under increased noise. The asterisks mark where the noise in training and evaluation match.

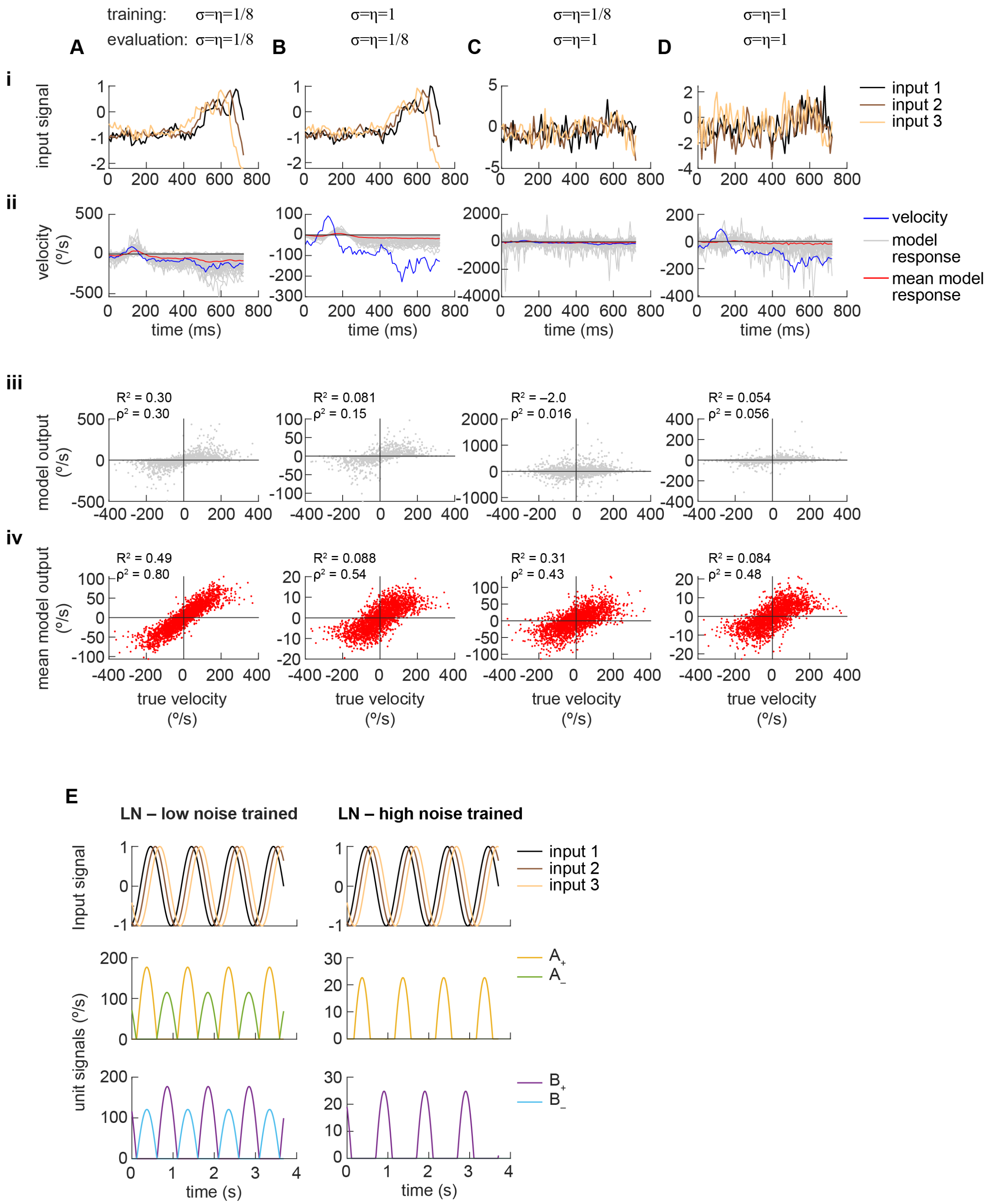

Figure S3. Natural scene responses in LN models trained under low and high noise conditions, related to Figure 4.

A) Training data responses for an LN model trained in low noise regime, responding to stimuli in low noise regime. (i) Time trace of example input signals. (ii) Time trace of true velocity and model predicted velocities. Blue line is true image velocity over time; gray lines correspond to models looking at different points in space; red line is mean of models over space. (iii) Scatter plot of instantaneous model output against true velocity. (iv) Scatter plot of mean model outputs, averaged over space, against true velocity. Variance explained (or coefficient of determination) is defined as $R^2 = 1 - \langle (v - \hat{v})^2 \rangle / \langle v^2 \rangle$, where v is the true velocity and $\hat{v}$ is the model output. The value ρ^2 is the square of the square of the Pearson correlation coefficient computed with v and $\hat{v}$. When the training and evaluation noise regimes are the same, $R^2 \approx \rho^2$.

B) As in (A), but with an LN model trained in high noise regime, tested in low noise regime. Note that the model does quite well, better than the situation in panel (C).

C) As in (A), but with an LN model trained in low noise regime, tested in high noise regime. Note how difficult the problem is with the added front end noise and back end multiplicative noise. The low noise trained model performs poorly here.

D) As in (A), but with an LN model trained in high noise regime, tested in high noise regime. Note that although the model is only performing modestly on this very difficult task, the average over space still does quite well up to a scaling factor, as indicated by the Pearson correlation. The model solution has a low amplitude and doesn't explain much variance by our metric, but yields results that are on average proportional to the velocity. Note that in high noise regimes, the multiplicative noise distribution at the back-end and the mean-squared-error loss function together favor average responses that are smaller than the true velocity, even when the multiplicative noise has an expected value of 1.

E) Responses of low- and high-noise trained LN model units to sinusoidal gratings moving in one direction. Note that in the low-noise trained model, the units are only mildly direction-selective (i.e., units A_+ and A_- respond with similar amplitude to this stimulus). In the high-noise trained model, the units are far more direction-selective, more similar to the biological detectors T4 and T5. This result motivated exploring the effects of noise on model solutions.