



OPEN

# Spatial attention in perceptual decision making as revealed by response-locked classification image analysis

Hironobu Sano, Natsuki Ueno, Hironori Maruyama &amp; Isamu Motoyoshi✉

In many situations, humans serially sample information from many locations in an image to make an appropriate decision about a visual target. Spatial attention and eye movements play a crucial role in this serial vision process. To investigate the effect of spatial attention in such dynamic decision making, we applied a classification image (CI) analysis locked to the observer's reaction time (RT). We asked human observers to detect as rapidly as possible a target whose contrast gradually increased on the left or right side of dynamic noise, with the presentation of a spatial cue. The analysis revealed a spatiotemporally biphasic profile of the CI which peaked at ~350 ms before the observer's response. We found that a valid cue presented at the target location shortened the RT and increased the overall amplitude of the CI, especially when the cue appeared 500–1250 ms before the observer's response. The results were quantitatively accounted for by a simple perceptual decision mechanism that accumulates the outputs of the spatiotemporal contrast detector, whose gain is increased by sustained attention to the cued location.

While humans and animals can immediately recognize scenes, objects, and materials at a glance<sup>1–4</sup>, in many situations they need to sequentially sample and integrate information to take more appropriate decisions about the world<sup>5,6</sup>. In this serial vision process, attention and eye movements play a critical role in the selection of relevant information from spatially distributed image inputs<sup>5,7,8</sup>. (It should also be noted that whether covert attention is parallel or serial has been controversial for decades<sup>9–12</sup>).

A large body of psychophysical evidence suggests that focal attention on the target's spatial location facilitates visual processing of the target<sup>6,13–20</sup>. In most of the studies, attention has been demonstrated to increase detection performance and accelerate the reaction to the target. However, such behavioral measures alone may be insufficient in explaining how attention affects the neural process of making decisions about the target, unless the results are compared across a variety of conditions.

In visual neuroscience, reverse correlation analysis has been widely applied to reveal information that determines the system response<sup>21,22</sup>. This analysis has been applied not only to the responses of cortical neurons<sup>23–25</sup> but also to human behavioral responses for a variety of visual tasks<sup>26–33</sup>. As a variant of the analysis, the classification image (CI) method allows the visualization of what information in the stimuli observers consider important for a given perceptual judgment<sup>34</sup>. In a typical experiment of the CI method, the observer's responses to a visual target embedded in white noise are collected, and the information in the stimulus that affected the observer's response is mapped out by analyzing the correlation between the noise and the response in each trial. The CI method has been widely used to reveal the spatiotemporal distribution of critical information (or the perceptive field) that determines the observers' judgments for various visual tasks with static and dynamic stimuli<sup>35–41</sup>.

Eckstein et al.<sup>36</sup> applied the CI method to Posner's cueing paradigm<sup>13</sup> and showed that the weight of information in the CI is greater at the spatial location where attention was directed<sup>36</sup>. However, observers in these studies made judgments after the visual stimuli had been shown, like in many psychophysical reverse-correlation studies<sup>28–30,32,33</sup>. Such post-stimulus judgment, usually based on the visual working memory, is not necessarily representative of the on-the-fly judgments that we make in real life. It is desirable to examine the effect of dynamic attention to the CI in the period before the observer responds to the shown target.

To clarify when observers make decisions during observation and what information they rely on to make decisions during observation, we can analyze correlations at each time point of the stimulus locked to the reaction time (RT) of the observer during the presentation of the stimulus rather than at the stimulus onset. Several

Department of Life Sciences, The University of Tokyo, Tokyo, Japan. ✉email: imotoyosi.ac@gmail.com

studies have adopted this response-locked reverse correlation analysis in investigating the dynamics of perceptual decision making<sup>27,42–44</sup>. Maruyama et al.<sup>45</sup> recently applied response-locked CI analysis to a very basic visual task, namely luminance contrast detection. Using a visual display similar to that used by Neri & Heeger<sup>37</sup>, they measured responses and reaction times for a target stimulus slowly emerging from dynamic noise and calculated the correlation between the noise and response at each time point, locked to the observer's reaction time (RT), in reverse chronological order. Adopting this protocol, they examined what signals and what point in the stimulus determined the observer's decision about the target and RT, and they revealed spatiotemporally biphasic CIs and their RT-dependent variability. They further suggested that the results could be quantitatively reproduced by a simple computational model incorporating a perceptual process approximated by a spatiotemporal filter and a (drift–diffusion) decision process that accumulates the output.

In the present study, we extended the above experimental protocol to investigate the effects of spatial attention. To this end, in addition to noise and target stimuli similar to those used by Maruyama et al.<sup>45</sup>, we presented a cue indicating the target's location (a valid cue) or another location (an invalid cue) as used in Posner's paradigm. Using this display, we measured responses and reaction times (RTs) for a target stimulus that appeared gradually in dynamic noise, and we compared RTs and CIs between the valid and invalid cue conditions. We found that the overall amplitude of temporally biphasic CIs was larger for the valid cue than for the invalid cue in the temporal period that the RT was shortened by attention. These results were well explained by a simple perceptual-decision model<sup>45</sup>, with the introduction of a single additional assumption that spatial attention sustainably increases the gain of the perceptual process.

## Methods

**Observers.** We initially recruited seven participants, which was the same number as in the previous study<sup>45</sup>. However, the data of one observer whose reaction time was found to be exceptionally long were excluded in the early phase of experiment. The results of the others showed that one of the expected effects—an interaction between the cue condition and the peak of the amplitude—was statistically significant. Based on the effect size in this result, we conducted an a priori power analysis using G\*Power 3.1 which revealed that a sample size of ten should be sufficient to achieve 0.95 power<sup>46</sup>. We then recruited four more participants, and finally ten human observers with corrected-to-normal vision, including three of the authors and seven naïve paid volunteers (having an average age of 24.2 years), participated in the experiment. All experiments were conducted in accordance with the Declaration of Helsinki. The study was approved by the Ethics Committee of The University of Tokyo. All observers gave written informed consent.

**Apparatus.** Visual stimuli were generated on a personal computer and displayed on liquid-crystal display monitors. To deal with the COVID-19 situation, the monitors (three BENQ XL2720B, three BENQ XL2730Z, one BENQ XL2735B, one BENQ XL2430T, one BENQ XL 2731K, one SONY PVM 2541A, and one SONY PVM-A250) were set up in the participants' own homes. The mean luminance of uniform backgrounds ranged from 44 to 116 cd/m<sup>2</sup>. The minimum luminance of the monitors ranged from 0.01 to 2.91 cd/m<sup>2</sup>, and the maximum luminance ranged from 87 to 232 cd/m<sup>2</sup>. All monitors had gamma-corrected luminance as calibrated with a colorimeter (ColorCal II CRS). The frame rate was 60 Hz. The viewing distance was adjusted to achieve a pixel resolution of 0.018 deg/pixel.

**Stimuli.** The visual display followed that of the previous study<sup>45</sup>. The stimulus was a pair of square images of dynamic one-dimensional noise (4.6 × 4.6 deg), each comprising 16 vertical bars with a width of 0.29 deg (Fig. 1). The contrast ( $C_{noise}(t)$ ) of each bar switched at a frame rate of 30 Hz according to a Gaussian distribution having a standard deviation of 0.1. The total stimulus duration was 4500 ms. Two independent one-dimensional noise fields were presented adjacent to the left and right of the gaze point.

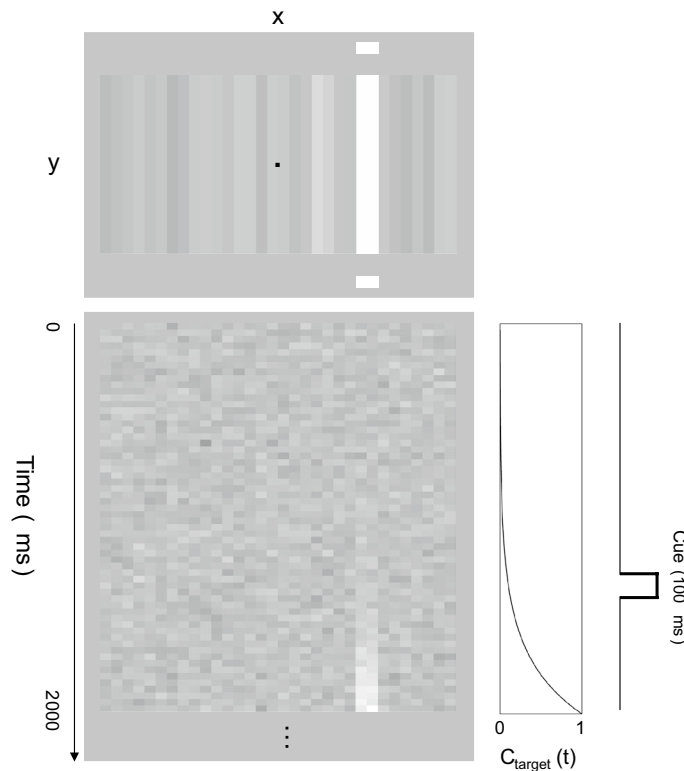
There are two target positions: left and right. The target signal ( $C_{target}(t)$ ) was added linearly to the two bars at the center of the noise area on either side. The addition of the target signal was set to start at a random frame between 0 and 500 ms after the start of the noise presentation. The luminance contrast of the target stimulus,  $C_{target}(t)$ , increased with time ( $t$ ) according to

$$C_{target}(t) = 10^{\min(0.05 \times t - 3, 0)}, \quad (1)$$

where  $t$  is the frame number (33 ms per unit) from stimulus onset and  $\min$  is the minimum function between  $0.05 \times t - 3$  and 0. The contrast of each bar was clipped in the range from  $-1$  to  $+1$ . The two fields, with and without the target signal, were transformed into a luminance image using the relation  $L(t) = L_{\text{mean}}(1 + C(t))$ , where  $L_{\text{mean}}$  is the mean luminance of the uniform background, which depended on the monitor of each observer (44–116 cd/m<sup>2</sup>).

In addition to the noise and target, a cue that probabilistically indicated the location of the target was presented. The cue was a pair of bright rectangles having a height of 0.3 deg and width of 0.6 deg. They were flashed for 100 ms at 0.6 deg above and below the central two bars in either the left or right field. The cue was presented at random timings of 500–2000 ms after the beginning of the increase in target contrast. The cue was presented in 94% of the trials. In 75% of these trials, the cue was presented at the target location (i.e., the valid cue condition), and in the remaining 25% of the trials, it was presented on the opposite side (i.e., the invalid cue condition).

**Procedure.** Observers binocularly viewed the display with a steady gaze at the fixation point and were asked to indicate by pressing a button whether the target appeared in the left or right field as rapidly as possible. Observers were informed that the cue would be presented at a valid location in the majority of trials. Audi-



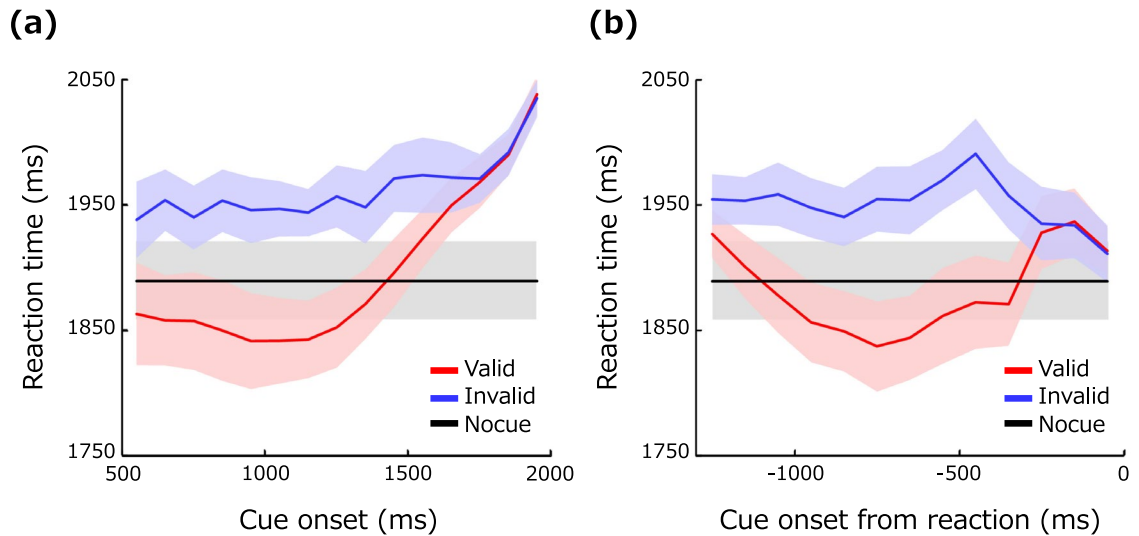
**Figure 1.** Schematic diagram of the visual stimuli used in the experiment. The upper panel shows a snapshot of a single stimulus frame. The bright bar on the right is the target, and the small horizontal bars above and below are the cue. The lower panel shows an X–T plot of the luminance variation of each bar. The target appears slowly in the right field. As shown in the right plot, the cue was presented for 100 ms with random timing from 500 to 2000 ms following the beginning of the increase in target contrast.

tory feedback was given for errors and responses that exceeded the deadline (4500 ms), and data in those trials were excluded from the analysis. The average error rate was 2%. The next trial started no less than 0.5 s after the observer's response. Each session of the experiment comprised 160 trials, and sessions were repeated until at least 3200 trials (4960 at maximum) were completed for each observer. In the analysis, all trials in which the observer responded before the cue, regardless of the validity of the cue, were included in the no-cue condition, along with trials in which the cue was not presented until the end.

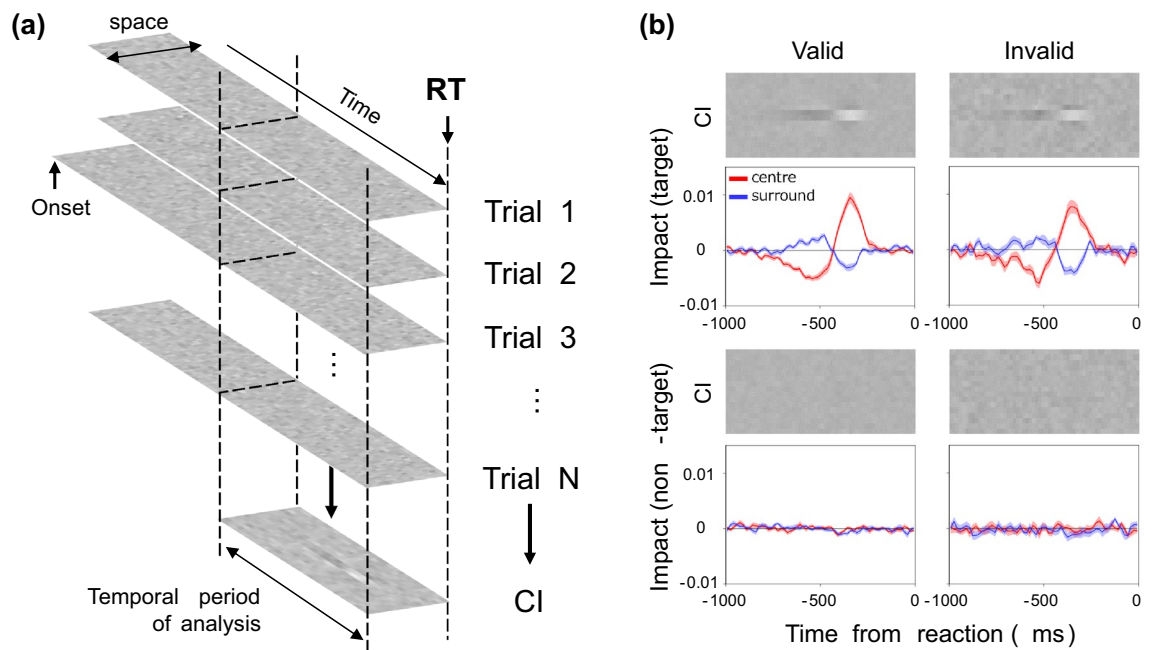
## Results

**Reaction time.** Figure 2a shows the reaction time (RT) as a function of the cue onset time. Each data point is the mean RT in each 100-ms epoch from the cue onset. The RT of the individual observer is defined as the harmonic mean, which is a more reliable index for measuring the central tendency of RT than the arithmetic mean<sup>47</sup>, across trials in each epoch. The red lines show the results for the valid cue, the blue line those for the invalid cue, and the black line those for no cue. RT is ~100 ms shorter for the valid cue than for the invalid cue, suggesting the effect of cueing. It is seen that the facilitation is weakened and diminishes when the cue is presented at ~1500 ms or later, but this is simply because this range contained slow-response trials in which observers responded after the contrast of the target had become high. A two-factor ANOVA revealed significant main effects of the cue condition ( $F(2,18) = 48.48, p < 0.0001, \eta^2 = 0.1011$ ) and cue onset time ( $F(14,126) = 33.36, p < 0.0001, \eta^2 = 0.0761$ ) and a significant interaction between them ( $F(28,252) = 21.18, p < 0.0001, \eta^2 = 0.0660$ ).

To better capture the dynamic variation in RT due to spatial cueing, we also calculated the mean RT as a function of the timing of the cue onset from the observer's response as shown in Fig. 2b. The plot shows that the RT was especially shortened when the valid cue was onset 500–1000 ms before the response. The RT was particularly long when the invalid cue was onset 500 ms before the response. We also find that the RT become longer when the cue was onset 1000 ms before the response, but this is simply because most of the responses in these trials were exceptionally slow regardless of the validity of the cue. It may appear strange that the difference in the RT between the valid and invalid cue conditions was small when the cue was onset ~300 ms before the response, but this difference was small simply because the cue was presented during the motor delay between the decision and button press (see also the section on computational modeling). A two-factor ANOVA revealed significant main effects of the cue condition ( $F(2,18) = 51.13, p < 0.0001, \eta^2 = 0.1074$ ) and cue onset from the response (-1500 - 0 ms) ( $F(14,126) = 15.05, p < 0.0001, \eta^2 = 0.0497$ ) and a significant interaction between them ( $F(28,252) = 18.41, p < 0.0001, \eta^2 = 0.0604$ ).



**Figure 2.** Effects of the spatial cue on the reaction time. **(a)** Reaction time as a function of the timing of the cue onset. Data are for each 100-ms epoch. The vertical axis shows the RT for the valid cue condition (red line), invalid cue condition (blue line), and no-cue condition (black line). The light-colored bands represent  $\pm 1$  S.E.M. across observers. **(b)** Reaction time as a function of the time from the response to cue onset. Data are for each 100-ms epoch.



**Figure 3.** **(a)** Diagram of the response-locked classification image. The classification image (CI) was calculated for each bar contrast at each spatiotemporal position backward from the observer’s response time. **(b)** The upper images show CIs for the valid (left) and invalid (right) conditions, where the vertical axis indicates space and the horizontal axis indicates the time from the response. The lower curves show the average weights of the two bars in the center (red curves) and the average weights of the two adjacent bars (blue curves). The light-colored bands represent  $\pm 1$  S.E.M. across observers. The upper panels show the results for the target field and the lower panels show the results for the non-target field.

**Classification image analysis.** Using the same procedure as in the previous study<sup>45</sup>, we conducted reverse correlation analysis of the contrast of each bar and the observer’s response (left, right) at each time point ( $t$ ) back from the reaction time, and we calculated the Classification Image (CI). Figure 3a is a diagram of the analysis. As in the work of Eckstein et al.<sup>36</sup>, CIs were calculated separately for the noise field where the target was presented and that where the target was not presented.

The images in Fig. 3b show the average CIs obtained for the target (upper panels) and non-target (lower panels) fields. For each, the left and right panels respectively show the results for valid and invalid cue conditions.

The brightness of each pixel represents the weight of noise on the response at spatial each position (vertical axis) and at each time before the response (horizontal axis). Bright pixels represent positive weights and dark pixels represent negative weights. The curves below the CI show the average weights of the two central bars (red) and the two adjacent bars (blue) in the CI. We refer to these as impact curves.

The CIs and impact curves obtained for the two cue conditions have similar trends, except that the amplitude appears slightly larger under the valid cue condition. For both cue conditions, there is clearly a characteristic spatiotemporal biphasic profile immediately before the response in the target field (upper panels). At the center of the CI where the target appeared, positive weights peak at approximately 350 ms before the response and negative weights at approximately 500 ms before the response. The opposite profile is found on both sides surrounding the center. These results, having clear agreement with the results of the previous study<sup>45</sup>, suggest that the observer's decision is triggered by the luminance increase after the luminance decrease in the central bars and the relative emphasis of these luminance variations in space. Meanwhile, the CIs of the non-target field (lower panels) show no systematic variation, and we will not consider these data in the following analysis (see also the discussion section regarding the interpretation of this result).

The RT data shown in Fig. 2b indicate that the observer's response was systematically accelerated when the valid cue was presented at a particular time before the response. With reference to these results on the RT, we next examine whether the CI varies between the valid and invalid cue conditions depending on how long before the response the cue appears. To this end, we divided the trials into epochs of 500 ms each based on the cue onset time from the response and calculated the CI for each epoch with a shift of 250 ms.

Figure 4a shows the CIs and impact curves for different times from the response to cue onset in the target field. The impact curves preserve a biphasic shape regardless of the cue onset time backward from the response. There is a negative weight of the center impact curve when the cue onset was long before the response. Given that trials in which the cue appeared long before the response are dominated by long RTs, this reflects the fact reported in the previous study that the negative weight is larger in trials of longer RTs<sup>45</sup>.

The overall amplitude of the impact curve is visibly different between the two cue conditions at a specific cue onset time backward from the response. Figure 4b plots the peak of the impact curve for each epoch. Here, the peak of the amplitude is defined as the average of the impacts within the range of 300–400 ms backward from the response. The red circles represent results for the valid cue condition and the blue circles those for the invalid cue condition. The solid line presents the results for the center impact and the dashed line those for the surrounds impact. We find that the positive weight is considerably larger for the valid cue condition than for the invalid cue condition when the cue was onset 500–1250 ms before the response, with the RT being especially shortened by the valid cue. This can be interpreted as the spatial attention being guided by the cue boosting the impact of information at that location.

When the cue was onset 0–500 ms before the response, we find that the difference between the two cue conditions is small, or the weights under the invalid cue condition appear to be slightly larger. This effect may appear to reflect suppressive effects of the cue onset itself on the response, but it is more likely a result of inclusion of trials in which the decision was made before the cue was presented, as was the case for the RT data (Fig. 2b).

A two-factor ANOVA was performed on the peak impact for the centre in Fig. 4b for the cue condition (valid, invalid) and the time from the response to the cue onset (five epochs). The results show that the main effect of the cue condition ( $F(1,9) = 4.40$ ,  $p = 0.0654$ ,  $\eta^2 = 0.0343$ ) was not significant, but the main effect of the time from the response to the cue onset ( $F(4,36) = 4.80$ ,  $p = 0.0033$ ,  $\eta^2 = 0.0711$ ) and the interaction ( $F(4,36) = 8.80$ ,  $p < 0.0001$ ,  $\eta^2 = 0.1060$ ) were significant. Corresponding t-tests conducted for each epoch showed significant differences for 0–500 ms ( $t(9) = -2.59$ ,  $p = 0.0291$ ,  $d = -0.6759$ ), 500–1000 ms ( $t(9) = 3.37$ ,  $p = 0.0082$ ,  $d = 1.0702$ ), and 750–1250 ms ( $t(9) = 3.04$ ,  $p = 0.0141$ ,  $d = 1.0031$ ). ANOVA on the peak impact for the surrounds showed no significant effect.

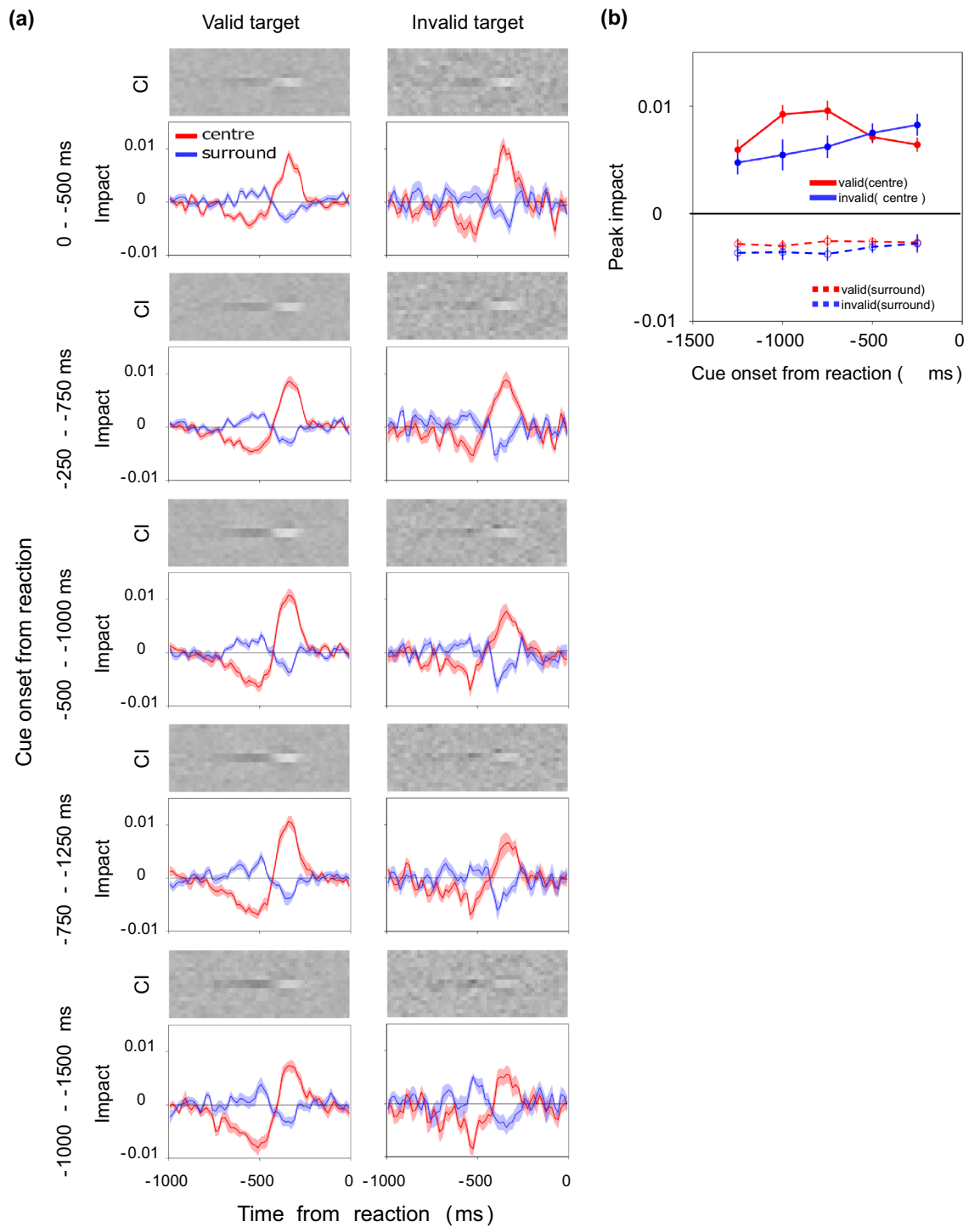
## Discussion

The present study applied the response-locked CI method to analyze the effect of attentional cueing on dynamic decision making in a simple contrast detection task. Consistent with the results of many previous studies<sup>13–15,48</sup>, we found that the RT was shortened by the valid cue, especially in the period of 300–1100 ms from cue onset. Furthermore, although the CIs had a spatiotemporally biphasic profile as in the previous study<sup>45</sup>, their overall amplitude was boosted by the valid cue, which also shortened the RT.

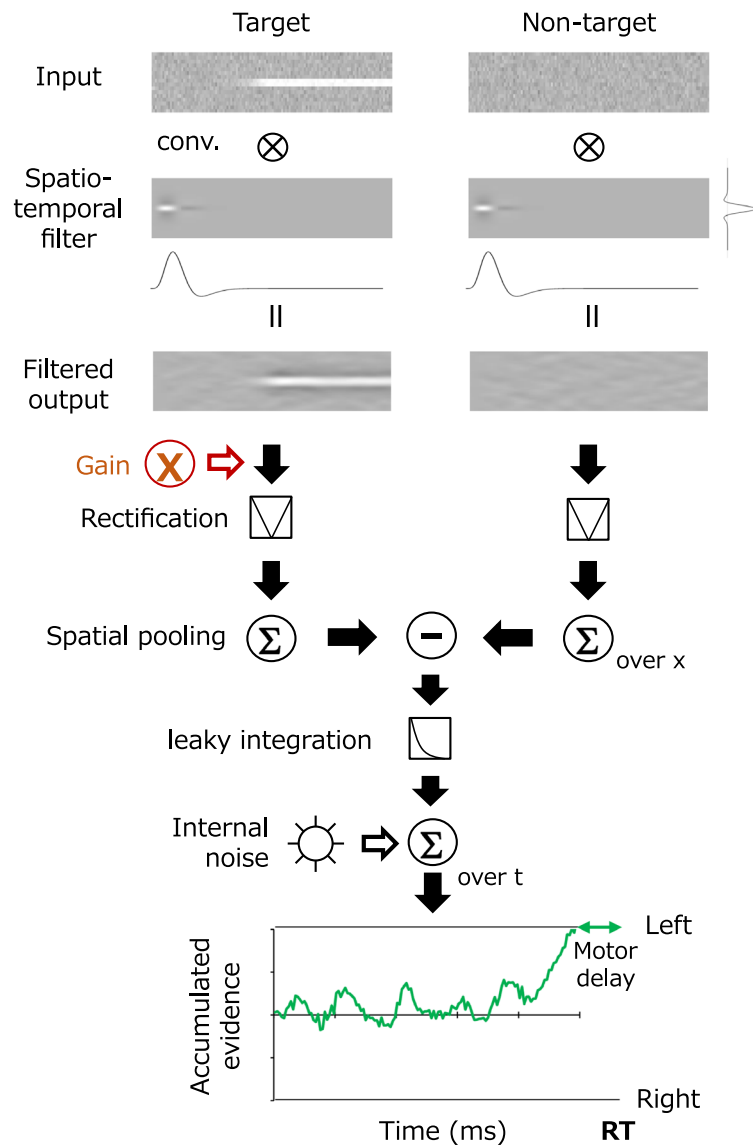
The reduction of the RT by the valid cue began to appear 300 ms from cue onset and was long lasting. Importantly, this response facilitation was observed even when the valid cue was presented more than 500 ms before the response. These results suggest that the valid cue summoned the observer's sustained attention to the target location and facilitated the response.

When cue the was onset 500–1250 ms before the response, CIs showed clear differences in the overall amplitude of the CI between the two cue conditions, while their spatiotemporal tuning profiles were relatively constant. Under the valid cue condition, the cue directed attention to the target region and increased the weight of information, thereby enhancing contrast detection sensitivity. Under the invalid cue condition, the cue directed attention to the opposite side of the target, which relatively de-emphasized information in the target region and suppressed contrast detection. These results are consistent with a large body of psychophysical evidence that indicate an increase in the contrast sensitivity to stimuli at an attended location<sup>18–20,49–52</sup> (see also the Computational model section regarding sensitization).

**Computational model.** On the basis of the above considerations, we examined whether the present results can be accounted for by a standard model of perceptual decision making with an additional assumption of the attentional enhancement of the perceptual sensitivity. The previous study assumed a hybrid model comprising a



**Figure 4.** (a) CIs and impact curves calculated for different times from the response to cue onset in the target field. From the upper to the lower panels, the results are for trials with the cue appearing shortly to long before the reaction. The left panels show the results for the valid cue condition and the right panels the results for the invalid cue condition. (b) Peak impact at approximately 350 ms before the response plotted as a function of the time from the response to cue onset. Red and blue circles represent results for the valid and invalid cue conditions, respectively. The solid and dashed lines show the results for the center and surrounds, respectively. Error bars are  $\pm 1$  S.E.M. across 10 observers.



**Figure 5.** Schematic diagram of a model based on spatiotemporal filtering and the accumulation of sensory evidence. See the main text for details.

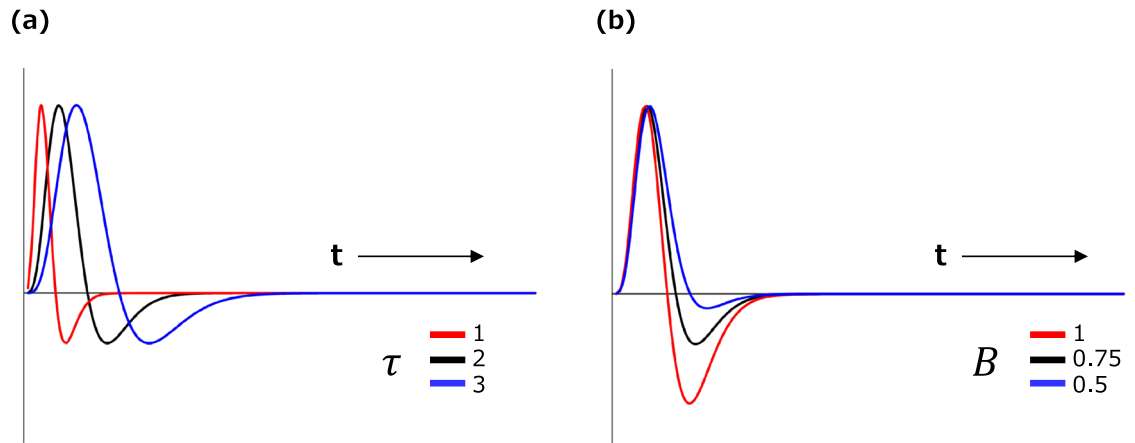
decision process that serially accumulates the outputs from the perceptual process, and it successfully predicted RTs and CIs in a similar display without spatial cueing<sup>45</sup>. Here, we extended the model by incorporating the effect of attention as an increase in the gain of the perceptual process, and we attempted to predict the characteristic variability of RTs and CIs depending on the cue validity.

Figure 5 presents an outline of the model. The model compares the spatially summarized outputs between two fields of the perceptual process approximated using a linear spatiotemporal filter. The decision process accumulates the differential signal of the two fields as sensory evidence over time and makes a decision when the evidence reaches a given boundary. The perceptual process is approximated as a spatiotemporal linear filter, and the decision process follows the standard drift-diffusion model for a two-alternative forced-choice task<sup>53–58</sup>. Figure 5 specifically illustrates each step in the case that the target appears on the left. In the model, the effect of spatial attention at the cued location is implemented as the amplification of the perceptual response in either the left or right field. The model assumed in the present study is very similar to traditional weighted integration models<sup>59,60</sup>. In particular, the effective sensitivity increase discussed as gain in our model can be interpreted as weighting in those models. The computation of each step is described in detail below.

The perceptual system is approximated as a space–time separable linear filter,  $F_{st}(x, t)$ , written as

$$F_{st}(x, t) = F_s(x) \cdot F_t(t), \quad (2)$$

where  $F_s(x)$  is the spatial filter,  $F_t(t)$  is the temporal filter,  $t$  is the frame number (33 ms per unit), and  $x$  is the pixel (0.018 deg per unit). The spatial filter  $F_s(x)$  is given as a difference-of-Gaussians function, which has been widely used as a first-order approximation for contrast detectors in the early visual system<sup>61</sup>.



**Figure 6.** Impulse response function of the temporal filter represented in Eq. 4. (a) Function from the transient factor ( $\tau$ ) from 1 to 3 ( $B = 0.75, n = 5$ ). (b) Function from the amplitude ratio between the positive and negative weights ( $B$ ) from 1 to 0.5. ( $\tau = 2, n = 5$ ).

$$F_s(x) = \exp\left(\frac{-x^2}{2\sigma_c^2}\right) - \frac{\sigma_c^2}{\sigma_s^2} \cdot \exp\left(\frac{-x^2}{2\sigma_s^2}\right). \quad (3)$$

Here,  $\sigma_c$  is the standard deviation of the center and  $\sigma_s$  is the standard deviation of the surrounds. The temporal filter  $F_t(t)$  is given as a biphasic function<sup>62,63</sup>:

$$F_t(t) = \left(\frac{1}{n!} - B \frac{(t/\tau)^2}{(n+2)!}\right) \cdot (t/\tau)^n \exp(-t/\tau), \quad (4)$$

where  $n$  is the number of stages of the leaky temporal integrator,  $\tau$  is the transient factor, and  $B$  is a parameter that defines the amplitude ratio between the positive and negative weights. Figure 6 shows the function of the temporal filter represented in Eq. (4).

The response of the perceptual system  $R(x, t)$  is then obtained by convolving the spatiotemporal filter  $F_{st}(x, t - t_0)$ , where  $t_0$  is the time of stimulus onset, with the stimulus input  $I(x, t)$ , and  $A$  is the gain of the system. Note that the gain  $A$  is assumed to be 1 typically and amplified in fields to which attention is directed by spatial cuing.

$$R(x, t) = A \cdot F_{st}(x, t - t_0) * I(x, t). \quad (5)$$

In the present study, we did not assume internal noise in the perceptual system for the practical purpose of limiting the number of model parameters; however, this omission might have resulted in ambiguities in the interpretation of the results. If the gain is implemented before the addition of location-specific internal noise, this results in an increase of the signal-to-noise ratio (sensitivity). By contrast, if internal noise is added after the integration of information, the gain becomes the weighting of evidence across locations without affecting the signal-to-noise ratio of individual locations. The difference between these two functions should be investigated by further experiment and modeling that focus on the question.

Determination of whether the target is presented in the left or right field is made by comparing the spatial sum of the absolute values of the responses in each field. The modeling is assumed to continuously monitor the difference  $\Delta R(t)$  between the left and right responses at time  $t$  from the stimulus onset. Here,  $\Delta R(t)$  is considered the sensory evidence at time  $t$  in the decision process:

$$\Delta R(t) = \sum_x |R_{\text{left}}(x, t)| - \sum_x |R_{\text{right}}(x, t)|. \quad (6)$$

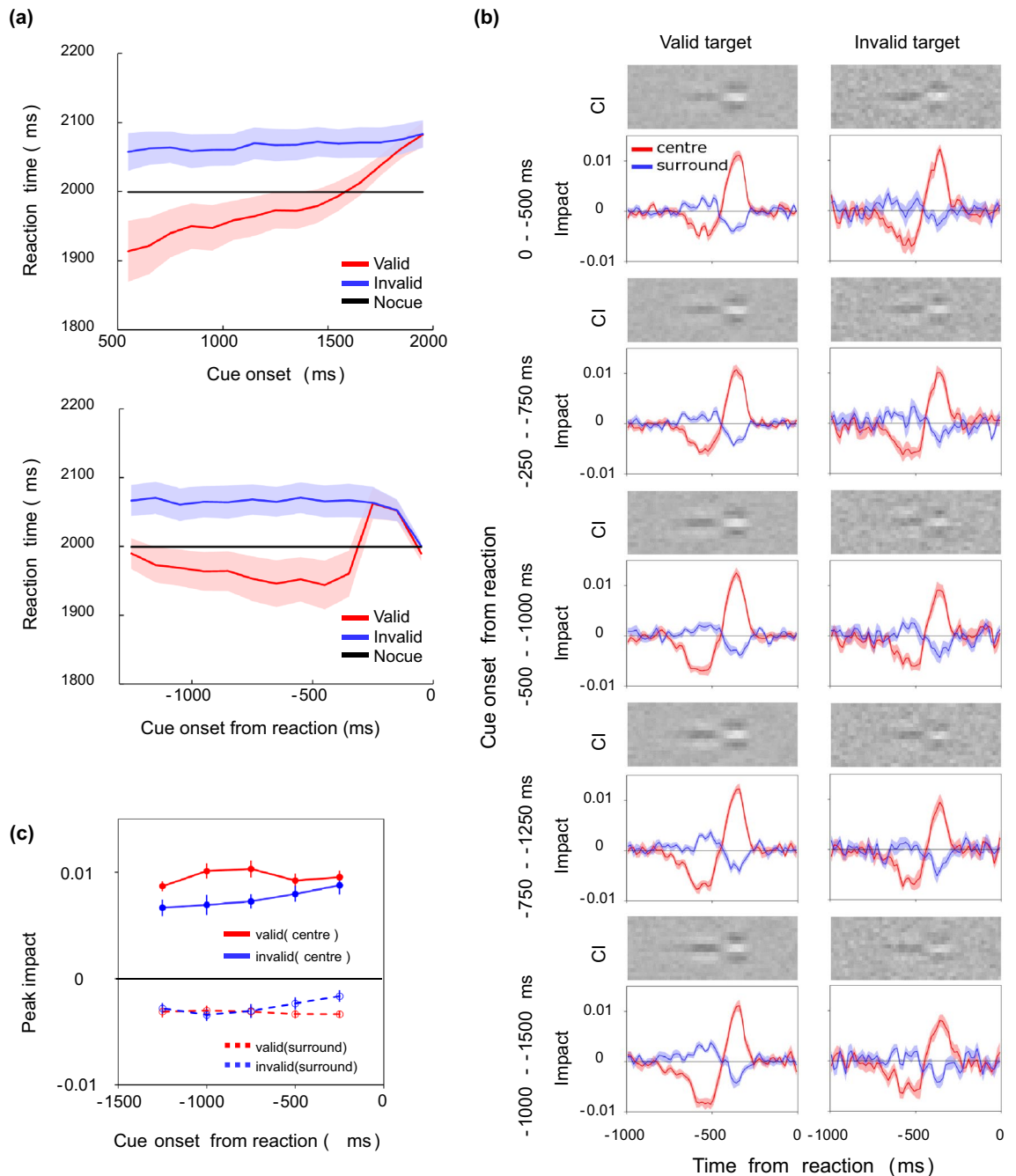
Decisions about targets are based on evidence accumulated over time. However, many decision-making studies suggest that sensory evidence decays with time<sup>29,33,64</sup>. This process is approximated by leaky temporal integration, which is mechanically ascribed to adaptive gain control<sup>65,66</sup>. The model thus assumes the cumulative evidence  $S(T)$  at time  $T$  is obtained from an approximation of the noisy leaky integral of  $\Delta R(t)$  as

$$S(T) = \sum_{t=1}^T (\gamma^{(T-t)} \Delta R(t) + \epsilon_t), \quad (7)$$

where  $\gamma$  is the time constant of evidence integration and  $\epsilon_t$  is the internal noise having a normal distribution. The model observer makes a decision about whether the target is on the left or right when  $S(T)$  reaches a certain decision boundary; i.e.,  $b$  or  $-b$ , respectively. Finally, the model observer is assumed to execute a manual response after a constant motion delay of 250 ms from  $T$ .

Our psychophysical data suggest a reduction in RT and an increase in the CI amplitude for the target at the cued location. According to the model architecture described above, this attentional facilitation must occur





**Figure 7.** Reaction times and CIs predicted by the model. Each panel corresponds to the results for human observers in Figs. 2 and 4. (a) Variation in the RT with cue. RTs for the no cue condition were estimated by averaging the valid and invalid conditions. (b) CIs and impact curves for each 500-ms epoch of the cue onset from the response. (c) Peak impact near 350 ms before the response.

prior to the comparison of the left and right fields. We therefore chose to incorporate the effect of attention as a sustained increase in the gain of the perceptual system following the cue onset. Specifically, we assumed that the increase in gain due to attention begins with a fixed delay of 50 ms following the cue onset.

In our model simulations, we attempted to reproduce the impact curve (Fig. 4) for the data of each observer, using the stimuli actually shown to the observer. The parameters of the model were estimated to minimize the RMS error with the impact curve, separately for each of the valid and invalid cue conditions. Under the invalid cue condition, the gain ( $A$ ) was fixed at 1. In accordance with previous studies<sup>45</sup>, the number of stages of the temporal filter ( $n$ ) was fixed at five.

Figure 7 shows the results of model simulation. It is seen that the model reproduces many aspects of the human data. Figure 7a shows a continuous reduction in the RT under the valid cue condition. Figure 7b and c

presents larger amplitudes of the CI under the valid cue condition than under the invalid cue condition, except for the period immediately after the cue onset.

Estimated parameters and the S.E. across model observers were  $[\sigma_c, \sigma_s, B, \tau, \gamma, b, \epsilon_t, A] = [2.94, 16.9, 0.505, 2.38, 0.207, 356.0, 10.6, 1.20]$  (S.E. = 0.13, 0.93, 0.006, 0.11, 0.004, 7.09, 0.69, 0.025) for the valid cue condition and  $[\sigma_c, \sigma_s, B, \tau, \gamma, b, \epsilon_t] = [2.92, 15.7, 0.490, 2.41, 0.212, 353.5, 9.80]$  (S.E. = 0.13, 0.59, 0.006, 0.10, 0.007, 7.22, 1.40) for the invalid cue condition. The relative gain  $A$  of the spatiotemporal filter (1.20) is obviously greater than 1 ( $t(9) = 7.69$ ,  $p < 0.0001$ ,  $d = 3.438$ ) for the valid cue condition, providing evidence of the sensory amplification by attention. Meanwhile, the other parameters did not significantly differ between the valid and invalid cue conditions, except for the ratio of positive to negative phases ( $B$ ) in the temporal filter, which was significantly larger under the valid cue condition ( $t(9) = 2.47$ ,  $p = 0.0355$ ,  $d = 0.8368$ ), but the difference was very small (0.505 vs. 0.490).

Taken together with the results of the previous study<sup>45</sup>, the present results suggest that perceptual decision making in simple contrast detection can be well described by a simple perceptual decision model. Adding to this basic result, the present data suggest that the facilitation of behavioral decision making via spatial attention is explained by the increased gain of the perceptual process. The effect of attention is not a modulation of the shape of the perceptual filter by the presence or absence of a cue but simply a change of gain (weighting), which is consistent with the results of previous studies<sup>36,60</sup>.

It is noteworthy that the variability of the RT due to a spatial cue was ~100 ms or more. This variability is considerably larger than values reported in many attention studies<sup>13,14,48</sup>. If we convert this 100-ms difference to the target bar contrast, we can infer that the valid cue enabled observers to detect targets with nearly half (~0.3 log units) the contrast of targets detectable with the invalid cue. One possible cause of this sensitization is the suppression of external noise by attention, which has been demonstrated by a larger attentional enhancement of contrast detection for a target flashed with noise than that without noise<sup>50</sup>. However, it has also been indicated that such attentional facilitation in noise is found only when the target is flashed together with noise<sup>50</sup>. The other possibility is the increase in the absolute sensitivity for targets with gradually increasing contrast over time, such as those used in the present study. Psychophysical evidence indicates that attention increases contrast sensitivity by 0.05–0.1 log units for targets of a stepwise waveform<sup>18,19</sup>, but by as much as 0.5–1.0 log units for targets with a gradual onset, regardless of the presence of external noise<sup>52</sup>. This large sensitization to the gradual target is observed even in the absence of background noise, which indicates an increase in the absolute detection sensitivity (i.e., reduction of internal noise) rather than the reduction of external noise<sup>52</sup>.

However, it is noted that the present finding does not rule out a variety of other possible explanations. Even assuming the model architecture shown in Fig. 5, it is difficult to determine exactly at what stage the attentional modulation of gain occurs. Our model assumes that the sensitivity of the spatiotemporal filter is modulated, but it is also possible that the spatially summarized output of the perceptual processes is modulated. Additionally, in the present study, we assumed that left and right evidence was integrated linearly; however, Shimozaki et al.<sup>60</sup> demonstrated that the sum of weighted likelihoods could more adequately approximate human observers. The magnitude of the gain (weighting) effect at each location could be modeled more elaborately. Furthermore, if we assume another type of decision model such as a ballistic model<sup>67,68</sup>, which can deal with two sensory evidence signals independently for each of the two target fields, one cannot discriminate whether attention affects the perception process for each target or the decision process for each evidence signal. As pointed out in previous studies<sup>45</sup>, it is generally difficult to distinguish between the involvements of perceptual and decision processes in the data of psychophysical reverse-correlation.

There are some limitations in the experimental design of the present study, which should be considered when interpreting the results. Initially, no result was obtained for CIs in the non-target field (Fig. 3b). This is likely to be caused by the design of the present experiment, in which observers could wait for the target until they were certain to respond to it. In fact, the average correct rate was 98%. There are two main problems with CI analysis that relies only on the target field. First, the target-present CIs are known to be biased in their estimation of the underlying weights the observer applies to the image<sup>69,70</sup>. Many different sets of weights of stimulus features (pixels) can result in a CI that looks like the target. Therefore, it is possible that the increase in impact (center) is partly attributed to the increase in luminance of the target. Second, as indicated in the literature<sup>70–72</sup>, uncertainty in the experiment could cause the difference in CIs between target and non-target fields. In the present study, it was essential that the target were gradual and presented continuously until it was detected by the observer. However, this inevitably resulted in an experimental paradigm with a large amount of temporal uncertainty.

Additionally, it should be noted that we could not distinguish between eye movements and covert attention shifts because eye tracking was not available. It has been suggested that the effects of visual search and cue can be explained by the same model<sup>36</sup>, and that the effect predicted by the model can be more fit to the observation with eye movement<sup>73</sup>. Although our observers were strongly instructed to maintain rigid fixation and the cue was flashed for a short time (100 ms), it was still possible for them to make eye movements during the stimulus presentation. Therefore, there should be caution regarding whether the present results are essentially ascribed to the effect of covert attention.

It is of interest to know how long the cueing effect persists. Murai and Whitney (2021) recently applied the CI method to the simple orientation detection task and demonstrated the serial dependence effect over several seconds<sup>74</sup>. Because the cue was rarely onset more than 1500 ms before the response in our experiment, it was difficult to systematically analyze the duration of the cueing effect. However, the CIs indicated that the cueing effect was persistent, even when the cue was onset 1000–1500 ms before the response (Figs. 4 and 7), which suggests that attentional facilitation could last longer than approximately 1 s. It would be worth exploring the possibility of long-lasting cueing effects in a future investigation using CI analysis.

## Data availability

The datasets generated and analyzed during the current study are not publicly available since they are concurrently used for a different project but are available from the corresponding author on reasonable request.

Received: 20 July 2022; Accepted: 17 November 2022

Published online: 05 December 2022

## References

1. Thorpe, S., Fize, D. & Marlot, C. Speed of processing in the human visual system. *Nature* **381**(6582), 520–522 (1996).
2. Oliva, A. & Torralba, A. Building the gist of a scene: The role of global image features in recognition. *Prog. Brain Res.* **155**, 23–36 (2006).
3. Motoyoshi, I., Nishida, S. Y., Sharan, L. & Adelson, E. H. Image statistics and the perception of surface qualities. *Nature* **447**(7141), 206–209 (2007).
4. Whitney, D. & Yamanashi Leib, A. Ensemble perception. *Annu. Rev. Psychol.* **69**, 105–129 (2018).
5. Treisman, A. M. & Gelade, G. A feature-integration theory of attention. *Cogn. Psychol.* **12**(1), 97–136 (1980).
6. Wolfe, J. M. Visual search in *The Handbook of Attention*, 27–56 (2015).
7. Duncan, J. Selective attention and the organization of visual information. *J. Exp. Psychol. Gen.* **113**(4), 501 (1984).
8. Treisman, A. & Gormican, S. Feature analysis in early vision: Evidence from search asymmetries. *Psychol. Rev.* **95**(1), 15 (1988).
9. Wolfe, J. M., Cave, K. R. & Franzel, S. L. Guided search: An alternative to the feature integration model for visual search. *J. Exp. Psychol. Hum. Percept. Perform.* **15**(3), 419 (1989).
10. Eckstein, M. P. The lower visual search efficiency for conjunctions is due to noise and not serial attentional processing. *Psychol. Sci.* **9**(2), 111–118 (1998).
11. Palmer, J., Ames, C. T. & Lindsey, D. T. Measuring the effect of attention on simple visual search. *J. Exp. Psychol. Hum. Percept. Perform.* **19**(1), 108 (1993).
12. Wolfe, J. M. Guided search 2.0 a revised model of visual search. *Psychon. Bull. Rev.* **1**(2), 202–238 (1994).
13. Posner, M. I. Orienting of attention. *Q. J. Exp. Psychol.* **32**, 3–25 (1980).
14. Posner, M. I. & Cohen, Y. Components of visual orienting. In *Attention and performance X: Control of Language Processes* 531–556 (Erlbaum, 1984).
15. Nakayama, K. & Mackeben, M. Sustained and transient components of focal visual attention. *Vis. Res.* **29**(11), 1631–1647 (1989).
16. Lee, D. K., Koch, C. & Braun, J. Spatial vision thresholds in the near absence of attention. *Vis. Res.* **37**, 2409–2418 (1997).
17. Lee, D. K., Itti, L., Koch, C. & Braun, J. Attention activates winner-take-all competition among visual filters. *Nat. Neurosci.* **2**, 375–381 (1999).
18. Carrasco, M., Penpeci-Talgar, C. & Eckstein, M. Spatial covert attention increases contrast sensitivity across the CSF: Support for signal enhancement. *Vis. Res.* **40**(10–12), 1203–1215 (2000).
19. Carrasco, M. Covert attention increases contrast sensitivity: Psychophysical, neurophysiological and neuroimaging studies. *Prog. Brain Res.* **154**, 33–70 (2006).
20. Carrasco, M. Visual attention: The past 25 years. *Vis. Res.* **51**(13), 1484–1525 (2011).
21. Neri, P., Parker, A. J. & Blakemore, C. Probing the human stereoscopic system with reverse correlation. *Nature* **401**(6754), 695–698 (1999).
22. Ringach, D. & Shapley, R. Reverse correlation in neurophysiology. *Cogn. Sci.* **28**(2), 147–166 (2004).
23. Ohzawa, I., DeAngelis, G. C. & Freeman, R. D. Stereoscopic depth discrimination in the visual cortex: Neurons ideally suited as disparity detectors. *Science* **249**(4972), 1037–1041 (1990).
24. DeAngelis, G. C., Ohzawa, I. & Freeman, R. D. Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. II. Linearity of temporal and spatial summation. *J. Neurophysiol.* **69**(4), 1118–1135 (1993).
25. Nishimoto, S., Ishida, T. & Ohzawa, I. Receptive field properties of neurons in the early visual cortex revealed by local spectral reverse correlation. *J. Neurosci.* **26**(12), 3269–3280 (2006).
26. Neri, P. & Levi, D. M. Receptive versus perceptive fields from the reverse-correlation viewpoint. *Vis. Res.* **46**(16), 2465–2474 (2006).
27. Ringach, D. L. Tuning of orientation detectors in human vision. *Vis. Res.* **38**(7), 963–972 (1998).
28. De Gardelle, V. & Summerfield, C. Robust averaging during perceptual judgment. *Proc. Natl. Acad. Sci.* **108**(32), 13341–13346 (2011).
29. Hanks, T. D. & Summerfield, C. Perceptual decision making in rodents, monkeys, and humans. *Neuron* **93**(1), 15–31 (2017).
30. Vandormael, H., Castañón, S. H., Balaguer, J., Li, V. & Summerfield, C. Robust sampling of decision information during perceptual choice. *Proc. Natl. Acad. Sci.* **114**(10), 2771–2776 (2017).
31. Sato, H. & Motoyoshi, I. Distinct strategies for estimating the temporal average of numerical and perceptual information. *Vis. Res.* **174**, 41–49 (2020).
32. Yashiro, R. & Motoyoshi, I. Peak-at-end rule: adaptive mechanism predicts time-dependent decision weighting. *Sci. Rep.* **10**(1), 1–8 (2020).
33. Yashiro, R., Sato, H., Oide, T. & Motoyoshi, I. Perception and decision mechanisms involved in average estimation of spatiotemporal ensembles. *Sci. Rep.* **10**(1), 1–10 (2020).
34. Ahumada, A. J. Jr. Perceptual classification images from Vernier acuity masked by noise. *Perception* **25**, 2 (1996).
35. Gold, J. M., Murray, R. F., Bennett, P. J. & Sekuler, A. B. Deriving behavioural receptive fields for visually completed contours. *Curr. Biol.* **10**(11), 663–666 (2000).
36. Eckstein, M. P., Shimozaki, S. S. & Abbey, C. K. The footprints of visual attention in the Posner cueing paradigm revealed by classification images. *J. Vis.* **2**(1), 3–3 (2002).
37. Neri, P. & Heeger, D. J. Spatiotemporal mechanisms for detecting and identifying image features in human vision. *Nat. Neurosci.* **5**(8), 812–816 (2002).
38. Mareschal, I., Dakin, S. C. & Bex, P. J. Dynamic properties of orientation discrimination assessed by using classification images. *Proc. Natl. Acad. Sci.* **103**(13), 5131–5136 (2006).
39. Rajashekar, U., Bovik, A. C. & Cormack, L. K. Visual search in noise: Revealing the influence of structural cues by gaze-contingent classification image analysis. *J. Vis.* **6**(4), 7–7 (2006).
40. Shimozaki, S. S., Chen, K. Y., Abbey, C. K. & Eckstein, M. P. The temporal dynamics of selective attention of the visual periphery as measured by classification images. *J. Vis.* **7**(12), 10–10 (2007).
41. Neri, P. & Levi, D. Temporal dynamics of directional selectivity in human vision. *J. Vis.* **8**(1), 22–22 (2008).
42. Busse, L., Katzner, S., Tillmann, C. & Treue, S. Effects of attention on perceptual direction tuning curves in the human visual system. *J. Vis.* **8**(9), 2–2 (2008).
43. Caspi, A., Beutter, B. R. & Eckstein, M. P. The time course of visual information accrual guiding eye movement decisions. *Proc. Natl. Acad. Sci.* **101**(35), 13086–13090 (2004).
44. Okazawa, G., Sha, L., Purcell, B. A. & Kiani, R. Psychophysical reverse correlation reflects both sensory and decision-making processes. *Nat. Commun.* **9**(1), 1–16 (2018).

45. Maruyama, H., Ueno, N. & Motoyoshi, I. Response-locked classification image analysis of perceptual decision making in contrast detection. *Sci. Rep.* **11**(1), 1–9 (2021).
46. Faul, F., Erdfelder, E., Lang, A. G. & Buchner, A. G\* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* **39**(2), 175–191 (2007).
47. Ratcliff, R. Methods for dealing with reaction time outliers. *Psychol. Bull.* **114**(3), 510–532 (1993).
48. Posner, M. I., Nissen, M. J. & Ogden, W. C. Attended and unattended processing modes: The role of set for spatial location. *Modes Perceiving Process. Inf.* **137**(158), 2 (1978).
49. Carrasco, M., Ling, S. & Read, S. Attention alters appearance. *Nat. Neurosci.* **7**(3), 308–313 (2004).
50. Lu, Z. L. & Doshier, B. A. External noise distinguishes attention mechanisms. *Vis. Res.* **38**(9), 1183–1198 (1998).
51. Ling, S. & Carrasco, M. Sustained and transient covert attention enhance the signal via different contrast response functions. *Vis. Res.* **46**(8–9), 1210–1220 (2006).
52. Motoyoshi, I. Attentional modulation of temporal contrast sensitivity in human vision. *PLoS ONE* **6**(4), e19303 (2011).
53. Ratcliff, R. A theory of memory retrieval. *Psychol. Rev.* **85**, 59 (1978).
54. Ratcliff, R. & Smith, P. L. A comparison of sequential sampling models for two-choice reaction time. *Psychol. Rev.* **111**, 333–367 (2004).
55. Huk, A. C. & Shadlen, M. N. Neural activity in macaque parietal cortex reflects temporal integration of visual motion signals during perceptual decision making. *J. Neurosci.* **25**(45), 10420–10436 (2005).
56. Gold, J. I. & Shadlen, M. N. The neural basis of decision making. *Annu. Rev. Neurosci.* **30**, 535–574 (2007).
57. Kiani, R., Hanks, T. D. & Shadlen, M. L. Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. *J. Neurosci.* **28**(12), 3017–3029 (2008).
58. Ratcliff, R. & McKoon, G. The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Comput.* **20**(4), 873–922 (2008).
59. Kinchla, R. A., Chen, Z. & Evert, D. Precue effects in visual search: Data or resource limited?. *Percept. Psychophys.* **57**(4), 441–450 (1995).
60. Shimozaki, S. S., Eckstein, M. P. & Abbey, C. K. Comparison of two weighted integration models for the cueing task: Linear and likelihood. *J. Vis.* **3**(3), 3–3 (2003).
61. Enroth-Cugell, C. & Robson, J. G. The contrast sensitivity of retinal ganglion cells of the cat. *J. physiol.* **187**(3), 517–552 (1966).
62. Adelson, E. H. & Bergen, J. R. Spatiotemporal energy models for the perception of motion. *Josa a* **2**(2), 284–299 (1985).
63. Watson, A. B. Temporal sensitivity. *Handb. Percept. Human Perform.* **1**(6), 1–43 (1986).
64. Usher, M. & McClelland, J. L. The time course of perceptual choice: The leaky, competing accumulator model. *Psychol. Rev.* **108**(3), 550 (2001).
65. Cheadle, S. *et al.* Adaptive gain control during human perceptual choice. *Neuron* **81**(6), 1429–1441 (2014).
66. Li, V., Michael, E., Balaguer, J., Castañón, S. H. & Summerfield, C. Gain control explains the effect of distraction in human perceptual, cognitive, and economic decision making. *Proc. Natl. Acad. Sci.* **115**(38), e8825–8834 (2018).
67. Brown, S. & Heathcote, A. A ballistic model of choice response time. *Psychol. Rev.* **112**, 117–128 (2005).
68. Brown, S. D. & Heathcote, A. The simplest complete model of choice response time: Linear ballistic accumulation. *Cogn. Psychol.* **57**, 153–178 (2008).
69. Ahumada, A. J. Classification image weights and internal noise level estimation. *J. Vis.* **2**(1), 8–8 (2002).
70. Tjan, B. S. & Nandy, A. S. Classification images with uncertainty. *J. Vis.* **6**(4), 8–8 (2006).
71. Barth, E., Beard, B. L. & Ahumada Jr, A. J. Nonlinear features in vernier acuity. In *Human Vision and Electronic Imaging IV*. **3644**, 88–96 (SPIE, 1999).
72. Solomon, J. A. Noise reveals visual mechanisms of detection and discrimination. *J. Vis.* **2**(1), 7–7 (2002).
73. Shimozaki, S. S., Schoonveld, W. & Eckstein, M. P. An ideal observer approach to unifying set size and cueing effects for perceptual and saccadic decisions. *J. Vis.* **5**(8), 514–514 (2005).
74. Murai, Y. & Whitney, D. Serial dependence revealed in history-dependent perceptual templates. *Curr Biol.* **31**(14), 3185–3191.e3 (2021).

## Acknowledgements

This study was supported by JSPS KAKENHI 20H01782.

## Author contributions

I.M. conceived the study. H.S., H.M., N.U., and I.M. designed the study. H.S., H.M., and N.U. performed experiments and analyzed the data. H.S. performed model simulation. H.S. and I.M. wrote the initial draft of the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to I.M.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022