1 **A response-locked classification image analysis of the perceptual**

2 **decision making : contrast detection**

3

4 Hironori Maruyama, Natsuki Ueno & Isamu Motoyoshi

5 Department of Life Sciences, The University of Tokyo

6 *(Submitted on April 10, 2021)*

7

## 8 Abstract

9 In many situations, humans make decisions based on serially sampled information

10 through the observation of visual stimuli. To quantify the critical information used by the

11 observer in such dynamic decision making, we here applied a classification image (CI)

12 analysis locked to the observer's reaction time (RT) in a simple detection task for a

13 luminance target that gradually appeared in dynamic noise. We found that the response-

14 locked CI shows a spatiotemporally biphasic weighting profile that peaked about 300 ms

15 before the response, but this profile substantially varied depending on RT; positive

16 weights dominated at short RTs and negative weights at long RTs. We show that these

17 diverse results are explained by a simple perceptual decision mechanism that

18 accumulates the output of the perceptual process as modelled by a spatiotemporal

19 contrast detector. We discuss possible applications and the limitations of the response-

20 locked CI analysis.

21

## 22 1. Introduction

23 While humans and animals can immediately recognize objects and scenes at a glance

24 (Thorpe, Fize & Marlot, 1996; Motoyoshi, Nishida, Sharan & Adelson, 2007; Whitney &

25 Yamanashi, 2018), in many situations they ensemble information in a sequence to take

26 more appropriate decisions (Bergen & Julesz 1983; Treisman & Gelade, 1980; Wolfe,

27 2015). In cognitive psychology, such dynamic information processing has been

1

28  investigated mainly by measuring the reaction times and correct rates of observers.

29  However, the reaction time alone is not powerful enough to reveal what kind of

30  information in the stimuli led the observers to make a decision at that moment in time,

31  unless data obtained under various conditions are compared.

32  In visual neuroscience, reverse correlation analysis is widely applied to reveal the

33  information in stimuli that determines the system responses (Neri & Levi, 2006; Neri,

34  Parker & Blakemore, 1999). This analysis has been applied not only to the responses of

35  cortical neurons (DeAngelis, Ohzawa & Freeman, 1993), but also to the analysis of the

36  behavioral responses of human observers (Ringach, 1998; Solomon, 2002). The

37  classification image (CI) method, one such technique, visualizes what information in the

38  stimuli observers consider important for a given perceptual judgement (Ahumada, 1996;

39  Abbey & Eckstein, 2002; Abbey & Eckstein, 2007). In typical experiments, the observer's

40  responses to a visual target embedded in white noise are collected, and the information in

41  the stimulus that affected the observer's response is mapped out by analyzing the

42  correlation between the noise and the response in each trial. The CI technique has been

43  used to reveal the spatial distribution of information, or perceptive field, that determines

44  the observer's judgments for a variety of visual tasks (Eckstein, Shimozaki & Abbey,

45  2002; Gold, Murray, Bennett & Sekuler, 2000; Rajashekar, Bovik & Cormack, 2006).

46  With dynamic stimuli, the CI method can also yield spatiotemporal perceptive fields (Neri

47  & Heeger, 2002; Mareschal, Dakin & Bex, 2006; Neri & Levi, 2008). Neri & Heeger

48  (2002) analyzed the correlation between spatiotemporal noise and responses in each trial

49  in a contrast detection task for luminance bars that slowly appear in dynamic noise. They

50  found CI profiles with biphasic weights in time and space, similar to the spatiotemporal

51  impulse response of the early visual system. Recently, similar psychophysical reverse

52  correlation with dynamic stimuli has been applied to the judgement on the average of

53  time-varying visual information to investigate the mechanisms of perceptual decision

54  making (Gardelle & Summerfield, 2011; Hanks & Summerfield, 2017; Li, Castañón,

55  Solomon, Vandormael & Summerfield, 2017; Vandormael, Castañón, Balaguer, Li &

56  Summerfield, 2017; Sato & Motoyoshi, 2020; Summerfield & Tsetsos, 2015; Yashiro &

57  Motoyoshi, 2020; Yashiro, Sato, Oide & Motoyoshi, 2020).

58    In the aforementioned studies, however, observers made decisions after the visual stimuli

59    had been shown. Such a judgment, which is usually based on visual working memory, is

60    somehow dissociated from on-the-fly judgments that we make in the real world. To

61    clarify when observers make decisions and what information observers rely on to make

62    decisions during observation, one can analyze correlations at each time point of the

63    stimulus locked to the reaction time of the observer during the presentation of the

64    stimulus rather than the stimulus onset. This response-locked reverse correlation has

65    been employed in several studies (Ringach, 1998; Busse, Katzner, Tillmann & Treue,

66    2008; Caspi, Beutter & Eckstein, 2004; Okazawa, Purcell & Kiani, 2018). For example,

67    Caspi et al. (2004) examined visual features that trigger saccadic eye movements by

68    analyzing the noise at time points locked to the onset of the saccade while the observers

69    views a multi-element display. Okazawa et al. (2018) adopted a reverse correlation

70    analysis locked to button-press responses to stochastic motion to explore the properties

71    of global-motion detectors and decision-making mechanisms.

72    In the present study, we applied the response-locked CI analysis to the most basic visual

73    task, luminance contrast detection. Specifically, we used stimuli similar to those used by

74    Neri & Heeger (2002) to measure responses and reaction times for target stimuli that

75    emerge slowly in dynamic noise, and we then analyzed the correlation between the noise

76    and response at each time point backward, locked to the observer's reaction time. This

77    protocol allowed us to examine what signals and what point in the stimulus determined

78    the observer's decision about the target and the observer's reaction time. The results

79    revealed spatiotemporally biphasic CIs similar to those reported by Neri & Heeger (2002).

80    On the other hand, we also found that the profile of the CI substantially varied depending

81    on the response time of the observer in a way that was unpredictable from the response

82    properties of the early visual system. These apparently complicated results, however,

83    were quantitatively described by a simple computational model incorporating a perceptual

84    process approximated by a spatiotemporal filter and a decision process (drift-diffusion)

85    that accumulates its output.

86

87    **2. Methods**

## 2.1. Observers

Five naïves and two of the authors (average age: 22.8 years) with corrected-to-normal vision participated in the experiment. All experiments were conducted with permission from the Ethics Committee of the University of Tokyo. Observers gave written informed consent. The study followed the Declaration of Helsinki guidelines.

## 2.2. Apparatus

Visual stimuli were displayed on a gamma-corrected LCD monitor (BENQ XL2735) controlled by a PC. The refresh rate was 60 Hz, and the pixel resolution was 0.04 deg/pixel at the viewing distance of 50 cm that we used. The mean luminance of the uniform background was 88.9 cd/m$^2$. All experiments were conducted in a dark room.

## 2.3. Stimuli

The visual stimulus was square dynamic one-dimensional noise (4.8 × 4.8 deg) comprising 16 vertical bars with a width of 0.3 deg (Fig. 1). The contrast ($C_{noise}(t)$) of each bar was switched at a frame rate of 30 Hz according to Gaussian noise with an RMS contrast of 0.1. The total duration was 8000 ms. Two independent 1D-noise fields were presented adjacent to the fixation point.
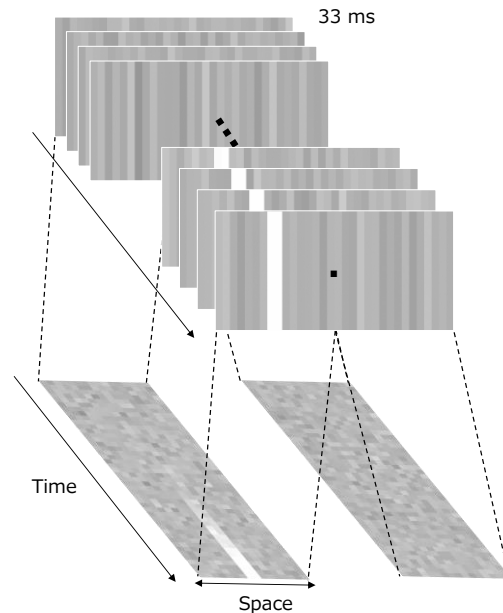
104

105     Figure 1. Schematic diagram of the visual stimuli used in the experiment. The bright bar that

106     appears on the left is the target.

107     The target signal ($C_{target}(t)$) was linearly added to the two central bars only for one of the

108     noise fields. The contrast of the target signal, $C_{target}(t)$, increased linearly with time ($t$) on

109     a logarithmic scale according to the following equation.

110     $\log C_{target}(t) = min(\alpha t - 3, 0)$                Eq. (1).

111     Here, $t$ is the time from the stimulus onset. $\alpha$ is the rate of increase, which was set at

112     three levels: 0.05, 0.1, and 0.2. The contrast of each bar was clipped in the range of $-1$

113     to $+1$. The two fields, with and without the target signal, were transformed into

114     luminance images using the relation $L(t) = L_{mean} (1+C(t))$, where $L_{mean}$ is the mean

115     luminance of the uniform background (88.9 cd/m$^2$).

116     **2.4. Procedure**

117     In each trial, observers viewed the stimulus at a fixation point binocularly and indicated

118     by pressing a button whether the target appeared in the left or right noise field as quickly

119     as possible. If an observer's response exceeded the deadline (8000 ms) or was an error,

120     auditory feedback was given, and the data recorded in that trial were excluded from the

5

121    analyses. The next trial started no less than 0.5 s after the observer's response. The

122    average error rates were 0.03, 0.02, and 0.02 for contrast increases ($\alpha$) of 0.05, 0.1, and

123    0.2, respectively. In each trial, the contrast values of all individual bars ($C_{noise}$ (x,t)), the

124    observer's response (left, right), and the reaction time were recorded. Each session of the

125    experiment comprised 150 trials for a single condition. For each observer, sessions were

126    repeated until at least 1200 trials were conducted for each condition.

127

## 128    3. Results

### 129    3.1. Reaction time

130    Figure 2a shows the average logarithmic reaction time of the observer, plotted as a

131    function of the contrast increase ($\alpha$). On a linear scale, the reaction times were 1949 ms

132    (s.e.= 36.1), 1207 ms (s.e.= 20.7), and 818 ms (s.e.= 22.7) for a contrast increase ($\alpha$)

133    of 0.05, 0.01 and 0.2, respectively. Fig. 2b is a cumulative histogram of the reaction

134    times of all observers. Fig. 2b shows that a slower rate of increase in the contrast

135    resulted in a longer average reaction time. One-way repeated-measure ANOVA on the

136    average reaction time with increasing contrast showed a significant effect of the rate of

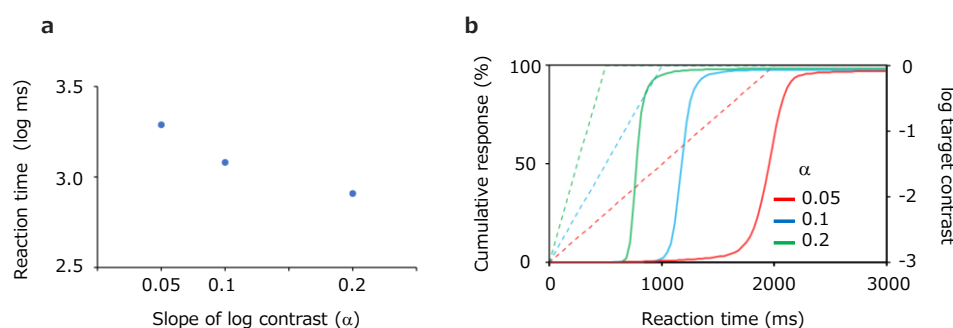137    contrast increase on the reaction time (F(2, 12) = 1574.6, p < 0.001).



138

139    Figure 2. (a) Average logarithmic reaction time as a function of the contrast increase rate. Error

140    bars represent ±s.e. across observers (invisibly small). (b) Cumulative histogram of reaction times

141    of all observers (solid line). Dashed lines represents the logarithmic contrast of the target stimulus

142    over time. Red, blue, and green lines represent a contrast increase of 0.05, 0.1, and 0.2,

143    respectively.

144 **3.2. Reverse correlation analysis locked to the response time**

145 We conducted a reverse correlation analysis between the contrast of each bar and the

146 observer's response (left, right) at each time ($t$) back from the reaction time to

147 characterize the noise common to the time before the reaction. Fig. 3 is a diagram of the

148 analysis. As in Neri & Heeger (2002), $\mu_1(x,t)$ is the mean of the noise contrast in the

149 region where the observer responded that the target was present and $\mu_0(x,t)$ is the mean

150 of the noise contrast in the region where the observer responded that the target was not

151 present. The results were calculated as follows.

152 $Mean\ Kernel\ (x,t) = \mu_1(x,t) - \mu_0(x,t)$         Eq. (2),

153 here Mean Kernel refers to the effect of the noise contrast on the response.
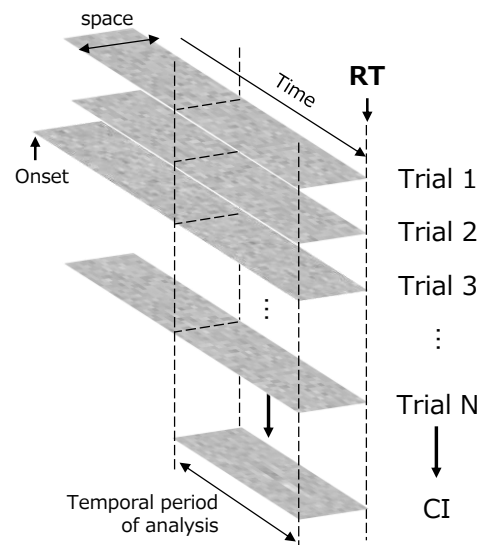


154

155 Figure 3. Reverse correlation analysis locked to the reaction. The classification image (CI) was

156 calculated for each bar contrast at each time from the reaction time.

157 The upper panels in Fig. 4 show the classification image (i.e., Mean Kernel) obtained in

158 the reverse correlation analysis locked to the reaction time. The horizontal axis represents

159 the time ($t$) back from the reaction time, and the vertical axis represents the spatial

160 position of each bar ($x$). In comparison with the grey background, the brighter points

161 represent positive weights and the darker points represent negative weights. Individual

162     panels show results for a contrast increase ($\alpha$) of 0.05, 0.1, or 0.2. The lower panels

163     show the mean of the weights of the two central bars (red) and the mean of the weights

164     of the two surrounding bars adjacent to them (blue) in the CI. The vertical axis

165     represents the weight and the horizontal axis represents the time from the reaction time.

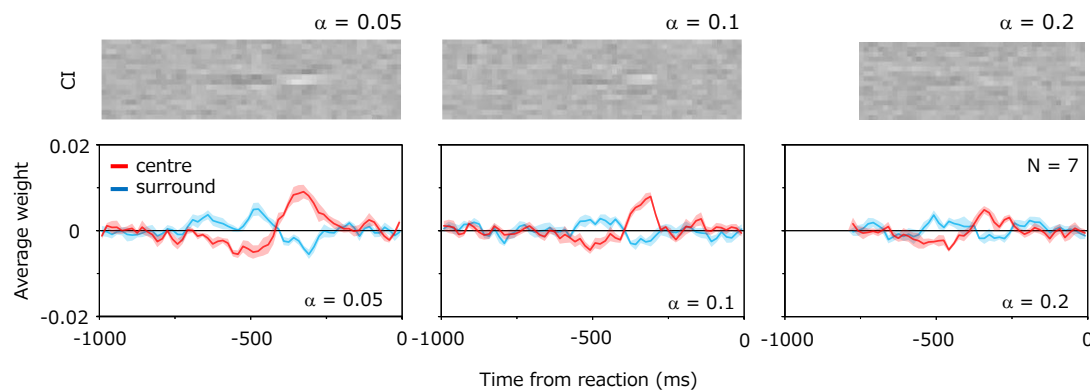166     We refer to the plots as impact curves.



167

168     Figure 4. Results of response-locked reverse correlation analysis. The upper panels show the

169     classification image (CI). The vertical axis represents the space and the horizontal axis represents

170     the time from the reaction time. Each pixel represents a positive (bright) or negative (dark) weight.

171     The lower panels show the average of the weights in the central two bars (red curves) and the

172     average of the weights in the two adjacent bars (blue curves). The vertical axis represents the

173     weights and the horizontal axis represents the time from the reaction time. Error bars represent

174     ±s.e. across observers. Individual panels show the results for a contrast increase ($\alpha$) of 0.05, 0.1,

175     or 0.2.

176     The above plots show characteristic temporal and spatial variations in the weights before

177     the target detection response. At the center of the stimulus where the target appeared, a

178     large positive weight was found about 300 ms before the response, and a negative

179     weight was found about 500 ms before the response. For the spatial variation, we find

180     that the weights around the target are reversed from the center. This indicates that the

181     central bar in which the target appears has high luminance compared with the

182     surrounding bars, which is a cue to the response. It is also found that the absolute

183     magnitude of the weights tends to increase as the contrast increase rate increases. We

184     also conducted the same analysis for contrast variance, as was done in a previous study

185    (Neri & Heeger, 2002), but found no clear CI profile at all.

186    For each condition, we calculated the average weights during the time epoch from −250
187    to −350 ms when in a clear variation in the weights is observed. We then conducted a
188    two-way repeated-measure ANOVA on the average weights with factors of the position
189    (center and periphery) and the contrast increase rate (a = 0.05, 0.1, 0.2) for each
190    condition. The results reveal that the main effect of position ($F(1,6) = 182.15$, $p < 0.001$)
191    was significant while the main effect of the contrast increase rate ($F(2,12) = 1.689$, $p =$
192    0.26) was not significant, and there was a significant interaction between the factors
193    ($F(2,12) = 20.28$, $p < 0.001$).

**3.3. Relationship between the RT and CI**

195    As shown in Fig. 2, the reaction time of the observer varied even under the same
196    conditions. Each individual observer responded quickly in some trials and took a long time
197    in others. Taking advantage of this fact, we investigated if and how the CI changes with
198    the reaction time. To this end, we divided the observer's data into 50% trials with short
199    reaction times, 50% trials with intermediate reaction times, and 50% trials with long
200    reaction times for each condition of the contrast increase rate, and carried out the reverse
201    correlation analysis for each group.

202    Figure 5 shows the CIs and impact curves obtained for each reaction time group. The
203    results were surprisingly different across the groups. The positive weights about 300 ms
204    before the reaction are larger for the shorter reaction time group. Conversely, the
205    negative weights 500 ms before the reaction are larger for the longer reaction time
206    group. This tendency is constant regardless of the contrast increase rate ($\alpha$). The results
207    indicate that the spatiotemporal profile of the weights of information correlated with the
208    response is remarkably different depending on the reaction time.
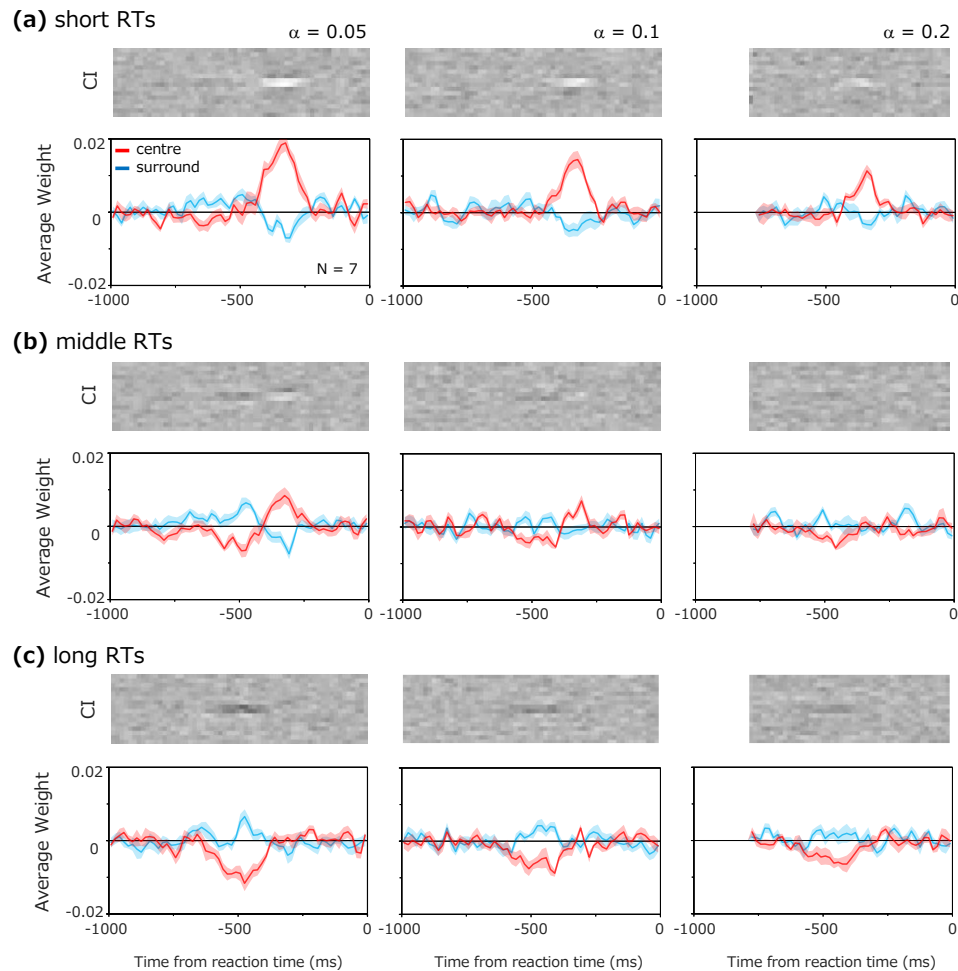
9

209

210     Figure 5. Results for each reaction time group. Results for (a) short reaction times, (b) intermediate

211     reaction times, and (c) long reaction times. Upper panels show the CIs and lower panels show the

212     impact curves.

213     For each condition, we calculated the average weights during the time epoch from −250

214     to −350 ms and during the time epoch from −450 to −550 ms. For each time epoch, we

215     performed a three-way-measure ANOVA on the average weights for the position (center

216     and periphery), the contrast increase rate (a = 0.05, 0.1, 0.2), and reaction time group.

217     For the time epoch of −250 to −350 ms, the results show that the main effects of the

218     position ($F_{(1,6)}$ = 269.03, $p < 0.001$) and reaction time group ($F_{(2,12)}$ = 28.44, $p <$

219     0.001) were both significant, and the main effect of the contrast increase rate ($F_{(2,12)}$ =

220     1.85, $p < 0.20$) was not significant. The results also show a significant interaction

221     between the position and reaction time group ($F_{(2,12)}$ = 54.025, $p < 0.001$), between

10

222 the contrast increase rate and reaction time group ($F_{(4,24)}$ = 4.85, $p < 0.01$), between

223 the position and contrast increase rate ($F_{(2,12)}$ = 18.18, $p < 0.001$), and between the

224 position and contrast increase rate and reaction time group ($F_{(4,24)}$ = 5.08, $p < 0.005$).

225 For the time epoch of $-450$ to $-550$ ms, the results show that the main effects of the

226 position ($F_{(1,6)}$ = 114.73, $p < 0.001$) and reaction time group ($F_{(2,12)}$ = 12.12, $p <$

227 0.005) were significant, and the main effect of the contrast increase rate ($F_{(2,12)}$ = 0.25,

228 $p < 0.784$) was not significant. The results also show a significant interaction between the

229 reaction time group and position ($F_{(2,12)}$ = 7.01, $p < 0.01$), between the reaction time

230 group and contrast increase rate ($F_{(4,24)}$ = 3.22, $p < 0.05$), and between the position

231 and contrast increase rate ($F_{(2,12)}$ = 5.90, $p < 0.05$). There was no significant

232 interaction between the position, contrast increase, and reaction time ($F_{(4,24)}$ = 0.750,

233 $p = 0.57$).

234

## 235 **4. Discussion**

236 The present study examined the information utilization strategy adopted in dynamic

237 decision making during stimulus observation in a simple contrast detection task. Applying

238 the classification image method, we calculated the weights of the embedded noise at

239 each time point retrospectively from the reaction time for the target. The resulting CIs

240 indicate that observers responded by utilizing the biphasic luminance change and the

241 central antagonistic spatial contrast before the response. In addition, we found that these

242 spatiotemporal profiles of CIs varied significantly depending on the reaction time.

243 The complex diversity of the results depending on the reaction time appears to be difficult

244 to understand intuitively. This may seem to indicate that the observers were so flexible

245 that they use different strategies for utilizing information depending on whether they

246 could respond quickly to the target or not. However, we should first consider a simple

247 explanation—that the results are a natural consequence of an interplay between the

248 sensory system and the decision process. Therefore, we tested a simple model consisting

249 of the early visual process (linear filtering model) and the perceptual decision process

250 (drift-diffusion model). As a result, we found that this conservative model with a fixed set

11

251    of parameters successfully duplicated the human data for all conditions and RT ranges.

**4.1. Computational model**

253    Figure 6 shows an outline of the model, which is inspired by a previous study on
254    spatiotemporal ensemble perception (Yashiro et al., 2020). The model compares the
255    spatially summarized outputs of the perceptual process, which is approximated using
256    linear spatiotemporal filters, between the two regions. The decision process accumulates
257    the differential signal between the two regions as sensory evidence over time and makes
258    a decision when the evidence reaches a given boundary. The basic structure of the
259    perceptual process follows that of a previous CI study (Neri & Hereger, 2002), and the
260    computation of the decision making follows traditional DDM modeling for a two-
261    alternative forced-choice task (Ratcliff & McKoon, 2008; Gold & Shadlen, 2007; Kiani,
262    Hanks & Shadlen, 2008). Fig. 6 shows each step of the process graphically for the case
263    that a target appears on the left. The calculation of each step is described in detail below.
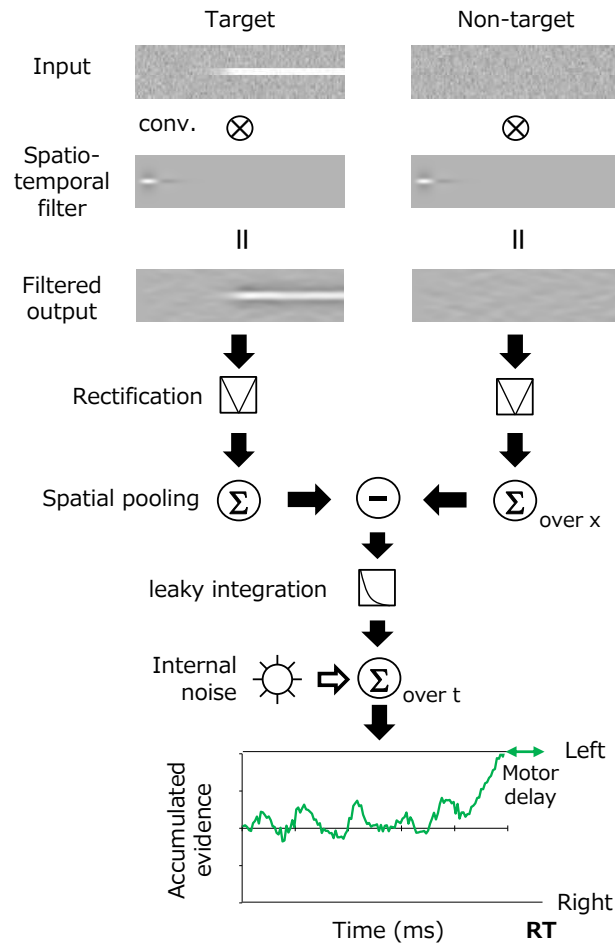
264

265   Figure 6. Schematic diagram of a model based on spatiotemporal filtering and the accumulation of

266   sensory evidence.


267   Following previous studies (Neri & Hereger, 2002), the perceptual system is

268   approximated as a linear spatiotemporal filter, $F_{st}(x,t)$, as follows.

269   $$F_{st}(x,t) = F_s(x) \cdot F_t(t) \qquad \text{Eq. (3)}.$$

270   Here, $F_s(x)$ is the spatial filter and $F_t(t)$ is the temporal filter. The spatial filter $F_s(x)$ is

271   given as a DoG function, which has been widely used as a first-order approximation for

272   contrast detectors in the visual system (Enroth-Cugell & Robson, 1966).

273   $$F_s(x) = exp\left(\frac{-x^2}{2\sigma_c^2}\right) - \frac{\sigma_c^2}{\sigma_s^2} \cdot exp\left(\frac{-x^2}{2\sigma_s^2}\right) \qquad \text{Eq. (4)}.$$

274   Here, $\sigma_c$ is the standard deviation for the central region and $\sigma_s$ is the standard deviation

13

275    for the surrounding region. The temporal filter $F_t(t)$ is given as the following biphasic

276    function (Watson, 1986; Adelson & Bergen, 1985).

277    $F_t(t) = \left(\frac{1}{n!} - B\frac{(t/\tau)^2}{(n+2)!}\right) \cdot (t/\tau)^n \exp(t/\tau)$        Eq. (5).

278    Here, $n$ is the number of stages in the time integrator, $\tau$ is the transient factor, and $B$ is a

279    parameter that defines the amplitude ratio of the positive and negative phases.

280    The response of the perceptual system was obtained by convolving the above

281    spatiotemporal filter $F_{st}(x,t)$ with the stimulus input $I(x,t)$.

282    $R(x,t) = I(x,t) * F_{st}(x,t)$            Eq. (6).

283    Decisions concerning whether the target presented in the left or right region were made

284    by comparing the spatial sum of the absolute values of the responses in each region

285    between the left and right. Thus, the model observer continually monitored the difference

286    $\Delta R(t)$ between the left and right responses at time $t$ from the stimulus onset. Here, $\Delta R(t)$

287    is regarded as the sensory evidence at time $t$ in the decision-making model.

288    $\Delta R(t) = \sum_x |R_{left}(x,t)| - \sum_x |R_{right}(x,t)|$        Eq. (7).

289    Decisions for targets are based on evidence accumulated over time. However, a number

290    of decision-making studies suggest that sensory evidence decays with time; that is, the

291    evidence weakens as it ages (Usher & McClelland, 2001; Hanks & Summerfield, 2017;

292    Yashiro et al., 2019). This property is practically described as a leaky temporal

293    integration, and it is potentially a product of the adaptive gain control of evidence signals

294    (Cheadle, Wyart, Tsetsos, Myers, Gardelle, Castañón & Summerfield, 2014; Li, Michael,

295    Balaguer, Castañón, Summerfield, 2018). According to these findings, the present

296    modeling assumes that the cumulative evidence $S(T)$ at time $T$ is given by the following

297    equation, which approximates the noisy leaky integration of $\Delta R(t)$.

298    $S(T) = \sum_{t=1}^{T}(\gamma^{(T-t)}\Delta R(t) + \epsilon_t)$        Eq. (8).

299    Here, $\gamma$ is the time constant of evidence integration and $\varepsilon_t$ is the internal noise following a

300    normal distribution. The model observer makes a decision about whether the target is on

301    the left or right when $S(T)$ exceeds a certain decision boundary, that is, $\beta$ or $-\beta$

302    respectively. The observer was assumed to execute a manual response after a constant

14

303    motor delay of 250 ms from $T$.

304    In this modeling, the perceptual process part has five parameters: the standard

305    deviations of the spatial filter ($\sigma_c$ and $\sigma_s$), the number of biphasic temporal filter

306    integration stages ($n$), the time constant ($\tau$) and the ratio of positive and negative phases

307    ($B$). The decision-making process part has three parameters: the decision boundary ($\beta$),

308    the internal noise ($\varepsilon_t$) and the time constant for evidence reduction ($\gamma$).

## 4.2. Model Simulation

310    We analyzed the CI and impact curves of the model observer using the image input data

311    that were presented to each observer in the experiment. In the simulations, for all data in

312    the condition of $\alpha = 0.05$, the model parameters were optimized for each observer to

313    minimize the squared error between the impact curve obtained for the model observer

314    and that of the human observer. To achieve the steady fitting, only the number of

315    integration steps of the biphasic temporal filter ($n$) was fixed to 5, for all model observers.

316    Figure 7 shows the simulation results. The thick impact curve represents the average of

317    results obtained for the optimized model for each observer, and the light-colored bands

318    represent the ±1 se range of the average for the human observer data. Estimated

319    parameters and the s.e. across model observers were [$\sigma_c$, $\sigma_s$, $B$, $\tau$, $\gamma$, $\beta$, $\varepsilon_t$] = [4.15,

320    17.4, 0.47, 0.793, 0.046, 496.5, 49.6] (s.e. = 0.18, 1.67, 0.07, 0.084, 0.002, 3.58,

321    2.64). For all values of the contrast increase ($\alpha$), we find that the model successfully

322    duplicated both the CI and the impact curve of the observers. For the three different

323    reaction time groups (Fig. 7b–d), the model duplicated the observers' data, reflecting the

324    characteristic differences of RT-dependent CI and impact curves. The root-mean-square

325    error (i.e., difference) between the fitted models and the model observers' data averaged

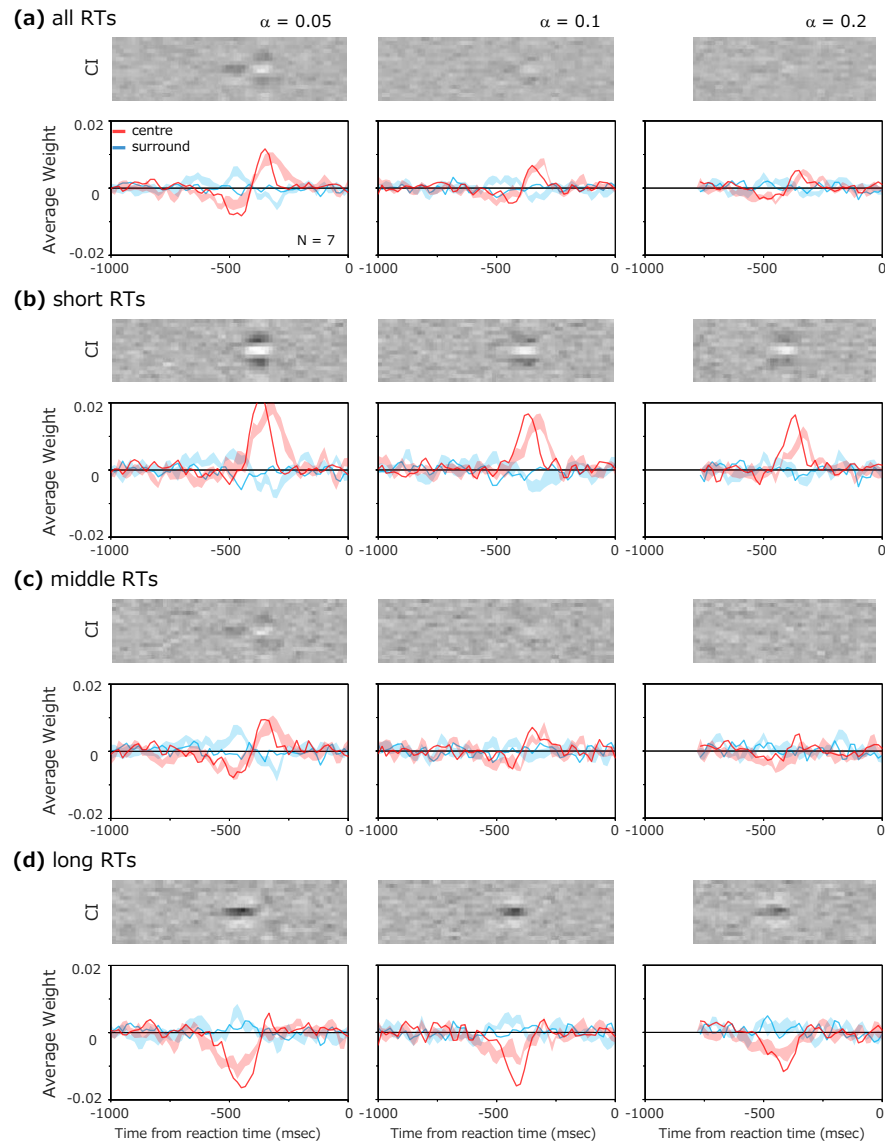326    over all observers was 0.005 (s.e. = 0.0001).

15

Figure 7. Results of the model simulations. (a) Impact curves and CIs derived from the data for all RTs. (b)–(d) Impact curves and CIs when divided into reaction time groups, with solid lines representing results for model observers and light-colored bands representing the ±1 s.e. range for human data.

To investigate the importance of the functional processes assumed in the model in Fig. 6, we simulated the model without some of the functional processes. We found that (1) the unique shape of the observers' CIs and impact curves could not be simulated if even one of the parameters of the spatiotemporal filter was omitted and (2) without the leaky

16

336    integration property being assumed in the decision making, the effect of the early stage

337    of stimulus presentation did not decrease even after a long observation in some RT

338    ranges. On the other hand, modifying the model to accumulate the responses in each

339    domain separately as two pieces of evidence and then calculate those differences, instead

340    of accumulating the differences in responses between the two domains as evidence, did

341    not change the behavior of the model, because the model essentially accumulates

342    evidence linearly (c.f., Bogacz, Brown, Moehlis, Holmes & Cohen, 2006).

343    The present results support the idea that on-the-view behavioral responses to visual

344    stimuli can be explained by a simple combination of the conventional perceptual model

345    and the standard perceptual decision-making model. This finding may allow us to perform

346    a response-locked reverse correlation analysis of human responses to sensory stimuli

347    during observation, rather than after observation, to explore the characteristics and

348    strategies of human information use in various cognitive tasks. In further investigations, a

349    similar framework may be used to understand the mechanisms for attentional selection

350    and for high-level visual cognition. The present computational model can be used as a

351    baseline account in these investigations.

352    It should be noted that psychophysical analysis cannot reliably separate the properties of

353    decision making from the low-level perceptual process (Okazawa et al., 2018). Although

354    one can partially overcome this limitation by making full use of various aspects of data,

355    such as by dividing the data into different RT ranges as in the present study, it is difficult

356    to distinguish between some properties such as the latency of the perceptual sensors and

357    the motor delay in the decision process.

358

### Acknowledgements

362

363

## References

1. Abbey, C. K., & Eckstein, M. P. (2002). Classification image analysis: Estimation and statistical inference for two-alternative forced-choice experiments. Journal of vision, 2(1), 5-5.

2. Abbey, C. K., & Eckstein, M. P. (2007). Classification images for simple detection and discrimination tasks in correlated noise. JOSA A, 24(12), B110-B124.

3. Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. Josa a, 2(2), 284-299.

4. Ahumada, A. J. Jr. Perceptual classification images from Vernier acuity masked by noise. Perception 25, 2 (1996).

5. Bergen, J. R., & Julesz, B. (1983). Parallel versus serial processing in rapid pattern discrimination. Nature, 303(5919), 696-698.

6. Bogacz, R., Brown, E., Moehlis, J., Holmes, P. & Cohen, J. D. The Physics of Optimal Decision Making : A Formal Analysis of Models of Performance in Two-Alternative Forced-Choice Tasks. Psychol. Rev. 113, 700–765 (2006).

7. Busse, L., Katzner, S., Tillmann, C., & Treue, S. (2008). Effects of attention on perceptual direction tuning curves in the human visual system. Journal of vision, 8(9), 2-2.

8. Caspi, A., Beutter, B. R., & Eckstein, M. P. (2004). The time course of visual information accrual guiding eye movement decisions. Proceedings of the National Academy of Sciences, 101(35), 13086-13090.

9. Cheadle, S., Wyart, V., Tsetsos, K., Myers, N., De Gardelle, V., Castañón, S. H., & Summerfield, C. (2014). Adaptive gain control during human perceptual choice. *Neuron*, *81*(6), 1429-1441.

10. DeAngelis, G. C., Ohzawa, I., & Freeman, R. D. (1993). Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. II. Linearity of temporal and spatial summation. Journal of Neurophysiology, 69(4), 1118-1135.

391  11. De Gardelle, V. & Summerfield, C. Robust averaging during perceptual judgment.
392      Proc. Natl. Acad. Sci. 108, 13341–13346 (2011).

393  12. Eckstein, M. P., Shimozaki, S. S., & Abbey, C. K. (2002). The footprints of visual
394      attention in the Posner cueing paradigm revealed by classification images The
395      Bayesian Observer : Cueing, 25–45.

396  13. Enroth-Cugell, C., & Robson, J. G. (1966). The contrast sensitivity of retinal ganglion
397      cells of the cat. The Journal of physiology, 187(3), 517-552.

398  14. Gold, J. M., Murray, R. F., Bennett, P. J., & Sekuler, A. B. (2000). Deriving
399      behavioural receptive fields for visually completed contours. Current Biology, 10(11),
400      663-666.

401  15. Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual*
402      *Review of Neuroscience*, *30*, 535-574.

403  16. Hanks, T. D. & Summerfield, C. Perceptual Decision Making in Rodents, Monkeys, and
404      Humans. Neuron 93, 15–31 (2017).

405  17. Huk, A. C. & Shadlen, M. N. Neural Activity in Macaque Parietal Cortex Reflects
406      Temporal Integration of Visual Motion Signals during Perceptual Decision Making. J.
407      Neurosci. 25, 10420–10436 (2005).

408  18. Kiani, R., Hanks, T. D., & Shadlen, M. L. (2008). Bounded integration in parietal
409      cortex underlies decisions even when viewing duration is dictated by the
410      environment. *Journal of Neuroscience, 28*(12), 3017–3029.

411  19. Li, V., Castañón, S. H., Solomon, J. A., Vandormael, H. & Summerfield, C. Robust
412      averaging protects decisions from noise in neural computations. PLoS Comput. Biol.
413      1–19 (2017).

414  20. Li, V., Michael, E., Balaguer, J., Castañón, S. H., & Summerfield, C. (2018). Gain
415      control explains the effect of distraction in human perceptual, cognitive, and
416      economic decision making. *Proceedings of the National Academy of*
417      *Sciences*, *115*(38), E8825-E8834.

418  21. Mareschal, I., Dakin, S. C., & Bex, P. J. (2006). Dynamic properties of orientation

419      discrimination assessed by using classification images. Proceedings of the National
420      Academy of Sciences, 103(13), 5131-5136.

421  22. Motoyoshi, I., Nishida, S. Y., Sharan, L., & Adelson, E. H. (2007). Image statistics
422      and the perception of surface qualities. *Nature*, *447*(7141), 206-209.

423  23. Neri, P., & Heeger, D. J. (2002). Spatiotemporal mechanisms for detecting and
424      identifying image features in human vision. Nature neuroscience, 5(8), 812-816.

425  24. Neri, P., & Levi, D. M. (2006). Receptive versus perceptive fields from the reverse-
426      correlation viewpoint. Vision research, 46(16), 2465-2474.

427  25. Neri, P., & Levi, D. (2008). Temporal dynamics of directional selectivity in human
428      vision. Journal of vision, 8(1), 22-22.

429  26. Neri, P., Parker, A. J., & Blakemore, C. (1999). Probing the human stereoscopic
430      system with reverse correlation. Nature, 401(6754), 695-698.

431  27. Okazawa, G., Sha, L., Purcell, B. A., & Kiani, R. (2018). Psychophysical reverse
432      correlation reflects both sensory and decision-making processes. Nature
433      communications, 9(1), 1-16.

434  28. Rajashekar, U., Bovik, A. C., & Cormack, L. K. (2006). Visual search in noise:
435      Revealing the influence of structural cues by gaze-contingent classification image
436      analysis. Journal of Vision, 6(4), 7-7.

437  29. Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for
438      two-choice decision tasks. *Neural Computation*, *20*(4), 873-922.

439  30. Ringach, D. L. (1998). Tuning of orientation detectors in human vision. Vision
440      research, 38(7), 963-972.

441  31. Sato, H., & Motoyoshi, I. (2020). Distinct strategies for estimating the temporal
442      average of numerical and perceptual information. Vision Research, 174, 41-49.

443  32. Shadlen, M. N. & Shohamy, D. Decision Making and Sequential Sampling from
444      Memory. Neuron 90, 927–939 (2016).

445  33. Solomon, J. A. (2002). Noise reveals visual mechanisms of detection and

446  discrimination. Journal of vision, 2(1), 7-7.

447 34. Summerfield, C., & Tsetsos, K. (2015). Do humans make good decisions? *Trends in*
448   *Cognitive Sciences, 19*(1), 27-34.

449 35. Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual
450   system. *Nature*, *381*(6582), 520-522.

451 36. Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention.
452   Cognitive psychology, 12(1), 97-136.

453 37. Usher, M. & McClelland, J. L. The time course of perceptual choice: The leaky,
454   competing accumulator model. Psychol. Rev. 108, 550–592 (2001).

455 38. Vandormael, H., Herce Castañón, S., Balaguer, J., Li, V. & Summerfield, C. Robust
456   sampling of decision information during perceptual choice. Proc. Natl. Acad. Sci. 114,
457   2771–2776 (2017).

458 39. Watson, A. B. (1986). Temporal sensitivity. Handbook of perception and human
459   performance, 1(6), 1-43.

460 40. Whitney, D., & Yamanashi Leib, A. (2018). Ensemble perception. *Annual Review of*
461   *Psychology*, *69*, 105-129.

462 41. Wolfe, J. M. (2015). Visual search.

463 42. Yashiro, R., & Motoyoshi, I. (2020). Peak-at-end rule: adaptive mechanism predicts
464   time-dependent decision weighting. Scientific Reports, 10(1), 1-8.

465 43. Yashiro, R., Sato, H., Oide, T., & Motoyoshi, I. (2020). Perception and decision
466   mechanisms involved in average estimation of spatiotemporal ensembles. Scientific
467   reports, 10(1), 1-10.

21