Spatiotemporal mapping of visual attention

Roger Koenig-Robert
CerCo, UPS, Université de Toulouse III, Toulouse, France, & CNRS, UMR5549, CHU Purpan, Toulouse, France

Rufin VanRullen
CerCo, UPS, Université de Toulouse III, Toulouse, France, & CNRS, UMR5549, CHU Purpan, Toulouse, France

The spatial distribution and the temporal dynamics of attention are well understood in isolation, but their interaction remains an open question. How does the shape of the attentional focus evolve over time? To answer this question, we measured spatiotemporal maps of endogenous and exogenous attention in humans (more than 140,000 trials in 23 subjects). We tested the visibility of a low-contrast target presented (50 ms) at different spatial distances and temporal delays from a cue in a noisy background. The cue was a non-informative salient peripheral (5°) stimulus for exogenous attention and a central arrow cue (valid 66.6%) pointing left or right for endogenous attention. As a measure of attention, we determined, for each distance and delay, the background contrast compensation required to keep performance at 75%. The spatiotemporal mapping of exogenous attention revealed a significant enhancement zone from 150 to 430 ms, extending up to 6° from the cue. Endogenous attention maps showed a peak at the cued side at 400 ms and between 8 and 10° from the cue. Modeling suggests that the data are compatible with a constant spotlight shape across time. Our results represent the first detailed spatiotemporal maps of both endogenous and exogenous attention.

Keywords: attention, detection/discrimination, contrast sensitivity


Introduction

Selective attention allows a sensory system to extract relevant information from the massive data influx reaching peripheral sensors. In order to do that, attention can use one or more of the following mechanisms: signal enhancement (i.e., amplitude increase of the task-relevant sensory inputs; Carrasco, Penpeci-Talgar, & Eckstein, 2000; Carrasco, Williams, & Yeshurun, 2002; C. C. Liu, Wolfgang, & Smith, 2009; Lu & Dosher, 2000), external noise exclusion (i.e., inhibition of non-relevant information; Dosher & Lu, 2000; Lu, Lesmes, & Dosher, 2002), and internal noise reduction (i.e., decrease in the variance of the perceptual processing; Mitchell, Sundberg, & Reynolds, 2003). Attention can also select relevant information by orienting or allocating the processing resources toward or away from a given signal source. In the visual domain, spatial orienting of attention can involve either eye movements toward a visual stimulus, which is known as overt spatial attention, or a selection of a part of the visual field without moving the eyes, known as covert spatial attention (Posner, 1980). A large body of behavioral evidence has dissociated two distinct forms of covert attention orienting: stimulus-driven orienting or exogenous attention and goal-driven orienting or endogenous attention (Cheal & Lyon, 1991; Chica & Lupiañez, 2009; Corbetta & Shulman, 2002; Egeth & Yantis, 1997; Hein, Rolke, & Ulrich, 2006; Nakayama & Mackeben, 1989; Yantis & Jonides, 1990; Yeshurun, Montagna, & Carrasco, 2008).

Exogenous attention is triggered by abrupt-onset or salient stimuli and is mostly reflexive (i.e., bottom-up) with little contribution from volitional states (Corbetta & Shulman, 2002; Egeth & Yantis, 1997; Giordano, McElree, & Carrasco, 2009; Yantis & Jonides, 1990). On the other hand, endogenous attention is triggered by higher level mechanisms (i.e., top-down) and depends on the subject’s expectations and knowledge (Corbetta & Shulman, 2002; Egeth & Yantis, 1997; Giordano et al., 2009; Hopfinger, Buonocore, & Mangun, 2000; Yantis & Jonides, 1990). Whether it is induced by exogenous or endogenous cues, the orientation of spatial attention toward a stimulus increases the perceptual performance (i.e., detection and/or discrimination) for stimuli falling within a neighboring spatial region (Eckstein, Peterson, Pham, & Droll, 2009; Egeth & Yantis, 1997; Huang & Dobkins, 2005; Lu & Dosher, 2000; Pastukhov, Fischer, & Braun, 2009; Posner, 1980; Shulman, Remington, & McLean, 1979; Talgar, Pelli, & Carrasco, 2004). Furthermore, this selection is not instantaneous but takes a certain time to develop and to subside (Müller & Rabbitt, 1989; Nakayama & Mackeben, 1989; Olivers & Meeter, 2008). These two attributes of visual attention, its spatial deployment pattern and its temporal dynamics, have often been studied in isolation. Over the years, a great amount of knowledge has been achieved about each of these two attributes independently.
Concerning the spatial deployment of attention, most data have come from studies in psychophysics (Cheal & Lyon, 1991; Eriksen & St. James, 1986; Posner, 1980; Sagi & Julesz, 1986; Shulman et al., 1979) but also from neuroimaging (Hopf et al., 2005; Kanwisher & Wojciulik, 2000; Reddy, Kanwisher, & VanRullen, 2009) and neurophysiology (Chelazzi, Duncan, Miller, & Desimone, 1998; Desimone & Duncan, 2000; Luck, Chelazzi, Hillyard, & Desimone, 1997; Spitzer, Desimone, & Moran, 1988). In these studies, the spatial pattern of attention is generally studied at a fixed time after the cue onset, corresponding to the expected peak of the attention effect (Posner, 1980; Sagi & Julesz, 1986). Thus, regarding its key spatial properties of filtering and enhancement, the focus of attention has either been modeled as a spotlight (Shulman et al., 1979), as a gradient (LaBerge & Brown, 1989), as a zoom lens (Eriksen & St. James, 1986; Shulman & Wilson, 1987), or as an inhibition beam (Tsotsos, 1990).

In the temporal domain, several studies of attention have characterized the time course of both exogenous and endogenous attention at a given spatial location—generally corresponding to the focal point of attention. The temporal dynamics of attention have been addressed in psychophysical studies (Fuller, Rodriguez, & Carrasco, 2008; Hein et al., 2006; T. Liu, Stevens, & Carrasco, 2007; Lyon, 1990; Mackeben & Nakayama, 1993; Müller & Rabbitt, 1989; Nakayama & Mackeben, 1989; Sagi & Julesz, 1986; Shimozaki, 2010; Theeuwes, Atchley, & Kramer, 2000; Weichselgartner & Sperling, 1987), in ERP studies, measuring the evoked electric responses in the brain (Clark & Hillyard, 1996; Harter, Miller, Price, LaLonde, & Keyes, 1989; Hillyard, Teder-Sälejärvi, & Münte, 1998; Mangun, 1995; Nobre, Sebestyen, & Miniussi, 2000), and also in single-unit recordings in non-human primates (Busse, Katzner, & Treue, 2008). These studies have revealed that attention peaks at about 80–150 ms after an exogenous cue, while the deployment of endogenous attention, triggered by a symbolic cue, takes about 300–500 ms and can be maintained at least for a few seconds (Ling & Carrasco, 2006b).

In most previous studies, however, the spatial pattern of attention has been implicitly thought to be fixed, only modulating its amplitude along the time domain. Hence, the temporal fluctuations of attention would be comparable regardless of the probed location, and its spatial extent would not depend on the exact probing time. Nevertheless, this assumption has rarely been tested directly (Kristjánsson & Nakayama, 2002; Tse, 2004). Using a change blindness-based paradigm, Tse succeeded in building up a map of exogenous attention deployment (over 149 positions) as a function of time (4 SOAs: 12, 82, 153, and 447 ms). Their principal result is an enhancing effect opposite the cued location, which, in some cases, was greater than the one at the cued location. Nevertheless, this work failed to show any clear modulation of attentional pattern across time. By and large, the interdependence between spatial and temporal dimensions of attention remains an unanswered question: Does the spatial focus of attention evolve along time? To address this question, we quantified the spatial deployment of attention as a function of post-cuing time in one single paradigm, for both exogenous and endogenous attention.

We tested the relative visibility of a low-contrast target presented at different distances (spatial sampling) and stimulus-onset asynchronies (SOAs) from the cue (temporal sampling) to build up a spatiotemporal map of attention effects, for both exogenously and endogenously driven attention. One original aspect of our paradigm is that we used a compensation technique to equalize the visibility of the target across screen positions, cue–target distances, and SOAs. In short, instead of adapting the intensity of the target to the various experimental conditions, we kept the target identical and varied the contrast of the noisy background using independent staircases for target eccentricity and several combinations of cue–target distance and cue–target SOA. The amount of background contrast modulation required to keep the target detection rate at 75% was taken as a reflection of the magnitude of attention at a particular distance and SOA from the cue (see Figure 1a and Methods section for more details). In our paradigm, the background noise contrast actually enhances the visibility of the target instead of interfering with its detection as in usual noise protocols (Lu et al., 2002; Talgar et al., 2004). This is because the target is only detectable if the background noise pattern has a sufficient contrast (the target is designed to be invisible when the background noise contrast is zero; see Figure 1a and Methods section). The use of noise in our paradigm does not imply that we favor the noise exclusion function of attention over its signal enhancement one. The conceptual framework underpinning our paradigm is that attention deployment enhances contrast sensitivity at the target location, a fact that has been very well documented (Cameron, Tai, & Carrasco, 2002; Carrasco et al., 2000; Ling & Carrasco, 2006a; Pestilli & Carrasco, 2005; Pestilli, Ling, & Carrasco, 2009). In order to test the dependency of the spatial deployment pattern across time, we compared our results to a model with null space–time interaction. The data for both exogenous and endogenous attention were well explained by the model, suggesting that the shape of the attention focus is constant over time (i.e., only its magnitude varies).

The development of this paradigm allowed us to obtain the first detailed spatiotemporal maps of both endogenous and exogenous visual attention. Moreover, the apparent
lack of reorganization of attention deployment pattern over time constrains the possible neural mechanisms that could underlie this cognitive function.

Methods

Subjects

All subjects were between 20 and 35 years old and had normal or corrected-to-normal visual acuity. They all gave informed consent. Thirteen subjects participated in the exogenous attention mapping and 10 subjects in the endogenous attention mapping.

Spatiotemporal mapping of attention

We used a detection task to quantify attention along the spatiotemporal domain. The target consisted of a gray cross (1°) that was present for 50 ms on 75% of trials, superimposed on a random background pattern of dark/light gray pixels, with the same mean luminance as the target (Figure 1a). The contrast of the background pixels was manipulated online and independently for each subject using a staircase method (Johnson, Chauhan, & Shapiro, 1992). Thus, the relative target visibility was modulated to equate detection performance at 75%: When the background contrast was zero, the target was invisible, as target and background had the same mean luminance; increasing the background contrast made the target more visible (Figure 1a). Thus, attention was measured as the contrast modulation of the background necessary to maintain performance at 75%.

In our paradigm, the background contrast (Cbg) was varied in space and time, modulated by 3 staircase-based functions, running simultaneously: a global contrast function (Cglob), an eccentricity function (Cecc), and a spatiotemporal attention function (Att) as follows:

$$C_{bg}(f, x, t) = (C_{glob} + C_{ecc}(f)) \cdot (1 + A_{tt}(x, t)). \quad (1)$$

The global contrast (Figure 1b) and eccentricity functions (Figure 2) provided the baseline for the actual attention measurement made using the spatiotemporal attention function (Figures 3 and 4). Note that the global contrast (Cglob) and eccentricity (Cecc) functions are in absolute units of contrast (between 0 and 1), while the spatiotemporal attentional (Att) function is only modulatory, with an expected value of 0 when attention has no effect.

The global contrast function compensated for intersubject variability, acting as a gain function. This function was a constant that modulated background contrast over

Figure 1. (a) Compensation paradigm. The background contrast is modulated in order to enhance or reduce target visibility, which in turn remains at the same luminance, fixing its detection rate at 75%. (b) An example of contrast compensation: Global contrast function defines the average contrast value all over the screen, fixing target detection rate at 75%. When the target is systematically detected, the background contrast decreases, making target detection harder (screen 2a). Inversely, when the target is systematically missed, the background contrast rises, making target detection easier (screen 2b). This adaptive staircase procedure allows maintaining the overall hit rate at 75%.
the entire screen area (Figure 1b). Contrast threshold was determined by one pair of staircases (one initiated at a low-contrast value and the other at a high-contrast value) in order to maintain performance at 75% no matter the eccentricity, the cue–target distance, or its SOA from cue onset.

The eccentricity function compensated for visual acuity differences between fovea and periphery. This was a function of distance from fixation point (d) and modulated the background contrast independently over 7 concentric segments centered at the center of the screen (Figure 2), updating its contrast depending on target eccentricity from fixation point in each trial. The contrast threshold for each segment was determined by a pair of staircases in order to maintain performance at 75% along all eccentricities.

Finally, the spatiotemporal attention function compensated for the effects of attention in both space and time. This was a function of distance (d) and delay (t) from cue, dynamically modulating the background contrast at different distances and delays from the cue (Figures 3 and 4). The screen area was divided in discrete segments centered at the cue position (spatial sampling) and trial duration was divided as well in discrete time steps (temporal sampling). The contrast threshold was determined independently using one staircase pair for each spatiotemporal coordinate, in order to maintain performance at 75% at every cue–target presentation distance and delay.

Each one of these staircase-based functions was updated on every target-present trial (75% of trials), converging to threshold value after a certain number of trials (Figure 5). The convergence criterion was defined as reaching a difference smaller than 1% contrast between the 2 thresholds in a given staircase pair. In this manner, the global contrast function converged to its threshold value quickly (in ~600 trials), defining the gain value. Then, the eccentricity function converged after ~1000 trials, making target visibility equal all over the screen no matter the eccentricity, except for the effects of attention that we aimed to determine. Finally, the attention spatiotemporal function (i.e., our measure of attention) converged to its threshold, making target equally visible no matter the distance and delay from cue.

**Exogenous attention mapping**

In exogenous attention mapping, we used a non-informative high-contrast peripheral cue flashed either for 66.6 ms (in the low-resolution condition, see below) or for 50 ms (in the high-resolution condition; Figure 3). The peripheral cue consisted of a dark square with white edges, measuring 0.2°. It was located randomly at 5° eccentricity. The target, a gray cross of 1°, was presented in 75% of trials. Since uniform target placement over the screen area would generate a bias (higher probability of appearing at higher eccentricities), we placed the target in a semi-random way: In each trial, we chose a random cue–target distance among those that had received the fewest number of trials so far.

We sampled the spatiotemporal domain by placing the target at 56 coordinates (7 distances × 8 time coordinates) in a low-resolution condition (6 subjects, 4,000 trials each). Cue–target distance was semi-randomly selected in each trial (as explained above) and then nested into 7 bins: 0.5, 1, 2, 4, 6, 8, 12°. Eight cue–target delays were probed: 0, 66.6, 133.3, 200, 266.6, 333.3, 400, 466.6 ms.

In a high-resolution condition (7 subjects, 10,000 trials each), we sampled the spatiotemporal domain by probing attention at 154 coordinates (14 distances × 11 delays). Cue–target distance was semi-randomly selected in each trial and then nested into 14 bins: 0.1, 0.5, 1, 1.5, 2, 2.5, 3, 4, 5, 6, 7, 8, 10, and 12°. Eleven cue–target SOAs were probed: 0, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500 ms.

**Target location control experiment**

We developed a control experiment to ensure that the semi-random target presentation did not provide information regarding the target position (i.e., the subject could
have guessed the position of the target). We restrained target presentation to the horizontal midline, thus avoiding the geometric bias for higher eccentricities. In this manner, we presented the target in a completely random way, positioning it uniformly over the screen midline. Four of the 13 subjects participated in this experiment (2 for the low-resolution condition and 2 for the high-resolution one).

**Endogenous attention mapping**

In endogenous attention mapping, we used an informative (66.7% of time) central cue pointing left or right. The cue was presented for the entire duration of the trial (700 ms). The target was presented on 75% of trials and flashed for 100 ms, placed randomly (uniform distribution) over the horizontal screen midline (Figure 4). Ten observers participated in this experiment.

We sampled the spatiotemporal domain at 70 coordinates (10 distances and 7 delays). Cue–target distance was randomly selected on each trial and then nested into 10 bins: −11, −7, −3, 0, 2, 4, 6, 8, 10 and 12°. The distance of 0° stands for the central cue position, positive distances define the cued side, and negative ones define the uncued side. Seven SOAs were probed: 0, 100, 200, 300, 400, 500, and 600 ms.
Stimulus generation

Stimuli were generated in Matlab, using the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997). Two distinct setups were used: one with a resolution of 800 × 600 pixels and a 100-Hz vertical refresh rate (corresponding to the high-resolution condition of the exogenous mapping experiment) and the other using a resolution of 1680 × 1050 pixels and a 60-Hz vertical refresh rate (corresponding to the low-resolution condition of the exogenous mapping experiment, as well as the endogenous mapping experiment). Absolute luminance was not calibrated, since the manipulated variable in our paradigm was relative luminance (i.e., contrast variation).

Procedure

Subjects were placed at 56 cm from the screen and were instructed to maintain fixation at the center of the screen. They were trained to detect the target (presented in 75% of trials) over a noisy background for a few practice trials before starting the experiment. They could respond by pressing two keys indicating target presence or absence.
Incorrect responses were fed back through a beep. Subjects had no time limitation to give their answer. A new trial began when response was given.

Exogenous attention mapping participants performed 4,000 trials each in the low-resolution condition (~4 h, across many sessions) and 10,000 trials each in the high-resolution condition (~10 h, across many sessions). Endogenous attention mapping participants performed 5,000 trials each (~5 h, along many sessions). Subjects received financial compensation for their participation.

Results

Exogenous attention mapping

First, we verified that our method was efficient at determining spatiotemporal attention effects by evaluating the convergence of the contrast thresholds for the attention function. The contrast for each spatiotemporal coordinate had been adjusted using two independent staircases (one randomly initialized at a low-contrast value and the other at a high value). We calculated the correlation (across spatiotemporal coordinates) between the estimated contrast thresholds for the attention (i.e., the background contrast must be diminished to compensate for the effect of attention and maintain performance at 75%; see Equation 1), while contrast modulations above the baseline indicate inhibitory effects of attention (i.e., the background contrast must be raised to compensate for the effect of attention). To visualize the spatiotemporal attention function, we interpolated (bicubic 2D interpolation) the original sampling grids (7 × 8 points for the low-resolution condition and 14 × 11 for the high-resolution condition) to a 100 × 100 point grid. One representative subject for the high-resolution condition is shown in Figure 6a.2. The X-axis corresponds to target delay from cue onset and the Y-axis to target distance from the cue. Attentional effects (percentage of background contrast compensation, see Equation 1) are color coded (from red to yellow representing attention enhancement, from blue to cyan representing inhibition, and black cor-
responding to the baseline). We then calculated the grand average across all subjects. Figure 7a shows the grand-average attention map for exogenous attention over 13 subjects. A progressive enhancement of attention can be seen from 50 to 150 ms. This enhancement attains a plateau from 200 to 350 ms, peaking at the spatial region extending up to 3° from the cue.

There is also a clear inhibitory effect starting at the origin (i.e., cue–target delay and distance = 0) and extending over 150 ms and 1°. This negative effect can be explained by the high contrast of the cue causing forward masking and inhibiting target detection (Enns & Di Lollo, 2000; Foley & Boynton, 1993), particularly at a short delay from the cue onset and at small distances from it. By definition, an optimal cue for capturing attention exogenously must be highly salient; thus, we believe that such forward masking was unavoidable in our paradigm.

In order to verify that the spatiotemporal attention pattern was reproducible across subjects, we calculated the Pearson correlation coefficient (across 100 × 100 spatiotemporal coordinates) between all possible pairs of subjects. The mean correlation coefficient across all 78 possible pairs ($C_{13}^2 = 78$) was $r = 0.2574$. We compared this result to the null hypothesis that the measured attention pattern was simply due to noise. To do this, we randomly scrambled the attention map of each subject (in its original 7 × 8 or 14 × 11 format, later interpolated over a 100 × 100 grid) and we calculated the Pearson correlation between all subject pairs as previously. This procedure was repeated 10,000 times and revealed that the probability of obtaining a correlation coefficient $r = 0.2574$ due to chance was less than $p < 10^{-4}$. Thus, the measured spatiotemporal attentional pattern is not simply due to noise but is consistent across subjects.
In order to further quantify this consistency of attention effects across subjects, we computed the significance of the main spatiotemporal region of attentional enhancement by means of a cluster analysis. (A cluster analysis takes into account the interdependence among the neighboring spatiotemporal coordinates and is thus more appropriate than classical parametric tests that assume independence of each point on the map.) Thus, we calculated the size of the cluster defined by all adjacent points above 3% contrast modulation in Figure 7a: The cluster size was 1695 points (out of 100 × 100 = 10,000 possible spatiotemporal coordinates). We then applied a bootstrap analysis to estimate the significance of this cluster size by comparing it with the null hypothesis that the same cluster size could be found due to chance. To achieve this, we shuffled the assignment of correct and incorrect responses to each trial for a given subject: The average performance would remain unchanged, but the delay and distance from cue would become meaningless variables. Thus, we recomputed contrast threshold value for every spatiotemporal coordinate and subject, obtained a grand-average attention map, and then reassessed the largest cluster size defined by all adjacent points above 3% contrast modulation—this time under the null hypothesis. We iterated this procedure 10⁴ times and found no cluster size equal or larger than 1695 points. Therefore, this bootstrap analysis reveals that attentional enhancement effects above 3% can be considered highly significant ($p < 10^{-4}$). These significant enhancing effects include the zone from 150 to 430 ms after cue onset and up to 6° of eccentricity.

To finish, we controlled that the semi-random cue position choice (see Methods section) was not used by subjects to anticipate the target position. We compared the control group of 4 subjects (where the target was located with a uniform random distribution over the screen horizontal midline) against the test group of 9 subjects (where at each trial a random location was chosen among those that had received the fewest number of trials so far). We calculated Pearson’s correlation coefficient between pairs of subjects belonging to the same group (i.e., control group or test group) or pairs belonging to different groups (one control subject with one test subject). The mean correlation into the same group was $r = 0.3193$ and across the groups, $r = 0.2430$. Bootstrap analysis (shuffling the assignment of subjects to the control or test group and repeating the pairwise correlation $n = 10^4$ times) showed no significant difference between control and test groups ($t$-test, $p = 0.104$). Thus, even if semi-random target placement could have provided information to anticipate the target position, test subjects did not appear to use it in a significant manner.

Endogenous attention mapping

As previously, we first evaluated the convergence of the contrast thresholds for the spatiotemporal attention function. We calculated the correlation (across spatiotemporal coordinates) between the estimated contrast thresholds of the 2 randomly initialized staircases and confirmed that this correlation increased throughout the experiment (Figure 5b, green line). In the same time, the correlation (across spatiotemporal coordinates) between a given staircase and its initial (randomly determined) value decreased steadily across trials (Figure 5b, dashed green line).
To visualize the spatiotemporal attention function, we interpolated (bicubic 2D interpolation) the original sampling grids (7 × 10 points) to a 100 × 100 point grid. Two representative subjects are shown in Figures 6b.1 and 6b.2. The X-axis corresponds to time from cue onset and the Y-axis to distance from the cue. Attentional effects are color-coded (red/yellow representing attention enhancement, blue/green representing inhibition, and black corresponding to the baseline). We then calculated the grand average across all subjects. Figure 8a shows the grand-average attention map for endogenous attention over 10 subjects. The average map shows an early (100 ms) enhancing effect centered on the cue, with a later deployment of attention at the cued side peaking between 8 and 10° at around 400 ms after cue onset.

In order to verify that the spatiotemporal attention pattern was reproducible across subjects, we calculated the Pearson correlation coefficient (across 100 × 100 spatiotemporal coordinates) between all possible pairs of subjects. The mean correlation coefficient across all 45 possible pairs ($C_{10^2} = 45$) was $r = 0.184$. We compared this result to the null hypothesis that the measured attention pattern was simply due to noise. To do this, we randomly scrambled the attention map of each subject (in its original 7 × 10, later interpolated over a 100 × 100 grid) and we calculated the Pearson correlation between all subject pairs as previously. This procedure was repeated 10,000 times and revealed that the probability of obtaining a correlation coefficient $r = 0.184$ due to chance was less than $p < 10^{-4}$. Thus, the measured spatiotemporal pattern of endogenous attention is not simply due to noise but is consistent across subjects.

In order to further quantify this consistency of attention effects across subjects, we computed the significance of the main spatiotemporal region of attentional enhancement by means of a cluster analysis, as previously. We calculated the size of the cluster defined by all adjacent points above 3% contrast modulation in Figure 8a: The cluster size was 929 points (out of 100 × 100 = 10,000 possible spatiotemporal coordinates). We then applied a bootstrap analysis to estimate the significance of this cluster size by comparing it with the null hypothesis that the measured spatiotemporal pattern was simply due to noise. To do this, we randomly scrambled the attention map of each subject (in its original 7 × 10, later interpolated over a 100 × 100 grid) and we calculated the Pearson correlation between all subject pairs as previously. This procedure was repeated 10,000 times and revealed that the probability of obtaining a correlation coefficient $r = 0.184$ due to chance was less than $p < 10^{-4}$. Thus, the measured spatiotemporal pattern of endogenous attention is not simply due to noise but is consistent across subjects.
The movies represent the experimental screen, centering the attention effects over the peripheral cue for exogenous attention mapping, and over the central cue for the endogenous one.

**Space–time interdependence**

Our compensation method was efficient at providing a detailed spatiotemporal map of exogenous and endogenous attention. One important question to address is whether the shape of the attentional focus changes over time from cue onset. To assess this, we compared our results to a null space–time interaction model. The modeled attention maps (Figure 7c for exogenous attention and Figure 8c for endogenous attention) were approximated as the linear combination of the spatial profile (i.e., the attention map averaged across time) and the temporal profile (i.e., the attention map averaged across space) as follows:

\[
\text{Att}_{(x,t)} = k(S_{(x)} \cdot M_{(t)}) + b,
\]

where the spatiotemporal attention function \(\text{Att}_{(x,t)}\) is decomposed in two parts: one spatial subfunction \(S_{(x)}\) that corresponds to the shape of the attentional focus and one activation function \(M_{(t)}\) that reflects the magnitude of attentional modulation as a function of time. The constant \(b\) represents baseline (i.e., average value of the spatiotemporal attention function), and \(k\) is a constant of proportionality obtained as follows:

\[
k = 1/ \int \int (\text{Att} - b)dx\,dt,
\]

defining \(S\) and \(M\) as

\[
S_{(x)} = \int (\text{Att} - b)dt,
\]

\[
M_{(t)} = \int (\text{Att} - b)dx.
\]

We calculated the modeled map for each subject and then subtracted it from the original map. The remainder of this subtraction is the attention effect that cannot be explained by a null space–time interaction. The residuals of the exogenous (Figure 7d) and endogenous (Figure 8d) mapping do not show significant values (cluster analysis, \(p = 0.8\) and \(p = 0.7\), respectively), suggesting that the spatial attention deployment pattern does not change significantly over time but only modulates its amplitude.

**Discussion**

**Spatiotemporal dynamics**

Our paradigm allowed us to estimate, for the first time, a detailed spatiotemporal map for exogenous and endogenous attentional effects, showing both enhancing and inhibitory effects.

The exogenous attention map reveals a significant enhancing effect zone from 150 to 430 ms, extending up to 6 deg from the cue. The enhancing effect is already visible (although not yet significant) at earlier latencies, rising progressively from 50 up to 150 ms. This slow and gradual rise could be due, in part, to forward masking by the cue over the target (Enns & Di Lollo, 2000; Foley & Boynton, 1993), inducing inhibitory effects that could cover up any earlier facilitatory attentional modulation. The enhancing effect in the exogenous attention map goes above 7% of modulation (yellow blobs, Figure 7a) and plateaus from 200 to 350 ms, in a spatial region extending up to 3 deg from the cue. This sustained effect could indicate a stabilization of the exogenous cuing, known to be triggered by predictive or interesting direct cues (Ward, 2008). The peak of the enhancement effect reaches values of about 10% contrast modulation. This means that, in this spatiotemporal facilitation zone, the target visibility is 10% higher than in the rest of the screen. A weaker inhibitory effect surrounds this zone, with a peak value of about −5% between 100 and 200 ms and reaching about −8% of modulation after 450 ms. However, analysis of individual subjects’ data revealed that this later inhibitory zone in the grand-average map was only due to two subjects who adopted a strategy of reporting the absence of the target whenever it appeared after 450 ms. None of the late inhibitory effects is suggestive of inhibition of return (IOR)—i.e., inhibition at previous attended locations. The absence of IOR can be explained by the fact that our exogenous attention map is limited in time to 466.6 ms, whereas IOR is known to appear with different onsets, ranging from 300 to 600 ms (Klein, 2000); thus, it is possible that we did not sample enough delays to reveal IOR in our paradigm. Previous reports have shown earlier modulations of transient attention using different paradigms. In a rapid serial visual presentation (RSVP) paradigm, Weichselgartner and Sperling (1987) measured the recall probability of the stream items occurring with and after the cue. They showed a cuing effect at latencies as early as 20 ms, with a decline at about 200 ms. In a discrimination task with a peripheral cue lasting 33 ms, Nakayama and Mackeben (1989) showed an enhancing peak in accuracy at about 100 ms, with a fast onset (from 0 to ~100 ms) and a slower decline (from ~100 to 500 ms). In a similar setting, Müller and Rabbitt (1989) revealed a peak of performance at 175 ms after the cue onset.
Theeuwes et al. (2000) showed an early facilitation of reaction times at 50 ms, followed by a drop of performance at 100 ms (due to the simultaneous presentation of a distractor), and then a stabilization of reaction times, with the fastest reaction times obtained from 150 to 300 ms. Attentional manipulations in the studies described above probably avoided forward masking by the cue at short SOAs and were thus more likely to obtain early attentional modulations.

Endogenous attention maps showed a peak at the cued side at ~400 ms and between 8 and 10° from the cue. The enhancing effect begins in the periphery (from 5 to 10°) and then spreads toward the cue location, reaching its maximum value of 12% at 10° from the cue at 420 ms. The existence of a second attention locus near the cue location (0°) is also visible as three different peaks (~200, ~350, and ~550 ms). This “second focus” could reveal that some subjects had trouble disengaging the exogenous component of their attention in some trials—thus enhancing target detection whenever it was presented near the cue. The endogenous attentional modulation time course showed in the present work is similar to previous reports using different paradigms. Weichelgartner and Sperling (1987) showed in an RSVP task that a symbolic cue instructing observers to switch the attended stream gated the deployment of attention at latencies of ~300 ms. Müller and Rabbitt (1989) showed that the time necessary to allocate spatial attention to the periphery after a valid central cue was of ~400 ms. In our mean attention map of both exogenous and endogenous attention, we obtained a peak of attentional modulation of around 10%, even though individual subjects often reached higher values (~20%, Figure 6). The modulation peaks across subjects do not necessarily coincide, which could explain the reduction in amplitude of the grand-average effect. The 20% peak of attention modulation from individual subjects falls into the range of amplitudes of attention modulation described in different physiological and psychophysical studies. Although behavioral and physiological data are not directly comparable (Eckstein et al., 2009; Pestilli et al., 2009), our effect seems to be smaller than the effect found in many single-unit reports in macaque monkey (30 to 40% gain in spontaneous activity in V2 and V4 in Luck et al., 1997; ~20% gain in neuronal responsiveness in V4 in Spitzer et al., 1988; 39 and 44% gain in firing rate in V4 and IT, respectively, in Chelazzi et al., 1998) or even in fMRI studies in humans (30% gain in BOLD signal in Reddy et al., 2009). Event-related potentials (ERPs) in humans (Kelly, Gomez-Ramirez, & Foxe, 2008) also reveal an enhancement of about 20% due to spatial attention in area V1. Overall, it could be that physiological response modulations (i.e., spiking rate, BOLD activity, and ERPs) are stronger than their behavioral counterpart. However, our main attention result is also in agreement with the gain found in the contrast response function in humans (Pestilli et al., 2009), where the difference in the threshold of the contrast response function (C50) between the cued and uncued conditions was less than 10%.

Method limitations

Several assumptions and approximations were necessary for our paradigm. For example, we defined the eccentricity function as a series of annular segments centered at fixation, thus assuming that visual acuity decreases homogeneously in both horizontal and vertical directions—an assumption we know to be overly simplistic (Smythies, 1996). Similarly, we sampled attentional effects in only one spatial dimension (i.e., the Euclidean cue–target distance), hence, assuming that attention is distributed radially from the cue location in a homogeneous manner. This assumption is not necessarily true since it has been shown that attention can be deployed preferably either over the vertical or the horizontal axis, depending on the visual context (Botta, Santangelo, Raffone, Lupiñez, & Belardinelli, 2010). We also assumed that attention maintains its deployment and time course characteristics regardless of whether the cue and target end up in the same or in different hemifields; yet interhemispheric deployment asymmetries and competition between hemifields have often been reported (Fecteau, Enns, & Kingstone, 2000; Mathôt, Hickey, & Theeuwes, 2010; Righi & Ribeiro-do-Valle, 2011). To counter the limitations of our paradigm, a mapping of attention in two spatial dimensions (i.e., in Cartesian x, y coordinates) instead of one (i.e., only Euclidean cue–target distance) would be necessary. However, the introduction of another spatial dimension would multiply the number of trials required and make the study impractical. All in all, our results can be taken as a good approximation to a spatiotemporal map of exogenous and endogenous attention, since many of the abovementioned undesirable manifestations of the spatial heterogeneity of attention will average out due to our counterbalancing the target presentation over all possible locations (i.e., the same cue–target distance but oriented in all directions, presenting cue and target in the same or in different hemifields, etc.).

Another issue is the degree of target location uncertainty in the exogenous versus endogenous condition. In fact, the possible target positions in the exogenous mapping includes all the possible positions on the screen (26 by 19.5°), and in the endogenous condition, the possible target locations includes all the possible positions within the horizontal meridian of 26° of length (13° at each side of the central cue), in both cases encouraging observers to select an extensive attended zone (i.e., a large attended field). Recent models (Reynolds & Heeger, 2009) and psychophysical data (Herrmann, Montaser-Kouhsari, Carrasco, & Heeger, 2010) have suggested that the amount of target uncertainty balances the attention effect
between contrast gain (a left shift in the psychometric curve with an attentional modulation in the intermediate contrast range) and response gain (an upward shift in the psychometric curve with an attentional modulation in the maximum contrast range). Contrast gain effect becomes predominant when the target size is significantly smaller than the zone of target uncertainty. Conversely, response gain becomes predominant when the target size is comparable or larger than the zone of uncertainty. In intermediate cases, the attention modulation becomes a combination of the two effects (contrast and response gain). In our paradigm, the zone of uncertainty is significantly larger than the target size in both exogenous (ratio between uncertainty area and target area equals 507) and endogenous mapping conditions (ratio equals 26). Critically, modeling predicts a predominant contrast gain effect for attention field/target size ratios equal or greater than 10 (Reynolds & Heeger, 2009). On the whole, previous evidence thus suggests that in our paradigm attention predominantly modulates contrast, validating our use of contrast as the dependent variable.

Our maps of exogenous attention, even in the “high-resolution” condition (data not shown) failed to confirm the existence of fine spatial features of attention: for example, we did not find a zone within the near surround of attention (2–3°) where attention declines to rise again at farther eccentricities, as suggested by Hopf et al. (2005) and Tsotsos (1990). This could mean that the signal-to-noise ratio in our paradigm was too low to detect the possible existence of subtle attention motifs.

Space–time independence

Our results suggest that the spatial pattern of both exogenous and endogenous attention does not change significantly across time; only its magnitude is modulated. In other words, the spotlight does not systematically shrink or expand as a function of time nor does it revert its polarity in any specific spatial subregion (e.g., from facilitation to inhibition). This result constrains the neural mechanisms that could support the deployment of visual attention. For example, it is compatible with a simple model composed of two independent neural populations. The first neural population, retinotopically organized, would encode the selected region of the visual field (the shape and the coordinates of the spatial attention pattern) in the form of a “saliency map,” as postulated by Itti and Koch (2000). The second neural source would represent the magnitude of attentional modulation as a function of time. In this model, the neural population that represents the spatial attention pattern (i.e., the saliency map) would receive global modulatory input from the second neural source, and there would thus be no dependence between the spatial pattern of attention and its magnitude modulation over time. Several putative anatomical localizations have been suggested for each of these two populations.

The equivalent of a “saliency map” for attention has been proposed to involve subcortical areas such as the pulvinar (Laberge & Buchsbaum, 1990; Robinson & Petersen, 1992), the lateral geniculate nucleus (Koch & Ullman, 1985), and the superior colliculus (Kustov & Robinson, 1996), as well as cortical locations such as V1 (Li, 2002), V4 (Mazer & Gallant, 2003), the frontal eye field (Thompson & Schall, 2000), and the posterior parietal cortex (Gottlieb, 2007). Neural populations that would gate the temporal modulations of attention have been identified by neuroimaging techniques in different regions such as the intraparietal sulcus, the lateral inferior premotor cortex, the cerebellum (Coull & Nobre, 1998), frontoparietal areas, and the thalamus (Fan, McCandliss, Fossella, Flombaum, & Posner, 2005; Posner, Sheese, Odludaş, & Tang, 2006). Note finally that the region providing the “temporal signal” for attention may not need to activate the saliency map with the precise time course that we have recorded here. Rather, one could also imagine that a simple “trigger” signal is sent to the saliency map. In this case, the intrinsic temporal properties of neuronal circuits within the saliency map would have to match the temporal modulation patterns observed in our experiments.

Acknowledgments

R.K. was funded by Conicyt and ANR, R.V. was funded by ANR (project 06JCJC-0154) and EURYI.

Commercial relationships: none.
Corresponding author: Roger Koenig-Robert.
Email: roger.koenig@cerco.ups-tlse.fr.
Address: 133 Rte de Narbonne, Toulouse 31062, France.

References


