

Chapter 19

The imaging of visual attention

Steven A. Hillyard¹, Francesco Di Russo,^{2,3} and
Antigona Martinez¹

1. Department of Neurosciences, UCSD, La Jolla, California □

2. Fondazione Santa Lucia IRCCS, Rome, Italy □

3. Istituto Universitario di Scienze Motorie (IUSM), Rome, Italy □

Abstract

Imaging studies reported in this book have made substantial advances in delineating the cortical networks responsible for allocating attention to locations and objects in the visual fields. Progress has also been made in specifying the multiple levels of the visual pathways at which incoming sensory information is selected or rejected. To reveal the dynamic mechanisms of attentional control and selection, it is necessary to employ imaging techniques having high temporal resolution (such as event-related potentials and magnetoencephalography) as well as techniques that offer anatomical precision (such as functional magnetic resonance imaging).

19.1 Introduction

Several contributors to this book have pointed out the fundamental distinction between the neural systems responsible for the control of selective attention and the sites within the sensory pathways where incoming information is actually modulated (Chapters 14, 15, and 17, this volume). Evidence from both neuroimaging and neuropsychological studies indicates that the control system for visual attention consists of a network of interconnected cortical regions that includes dorso-lateral frontal and posterior parietal areas. It is hypothesized that anatomical projections run from this control system to appropriate levels of the visual pathway and exert facilitatory and/or inhibitory influences on the neurons that encode incoming information. One version of this general scheme is illustrated in Fig. 19.1, where the red arrows represent attentional control signals and the blue arrows the flow of sensory information.

19.2 Levels of selection in spatial attention

A key research goal within this framework has been to identify the level(s) of the visual pathways where sensory inputs are modulated by the attentional control

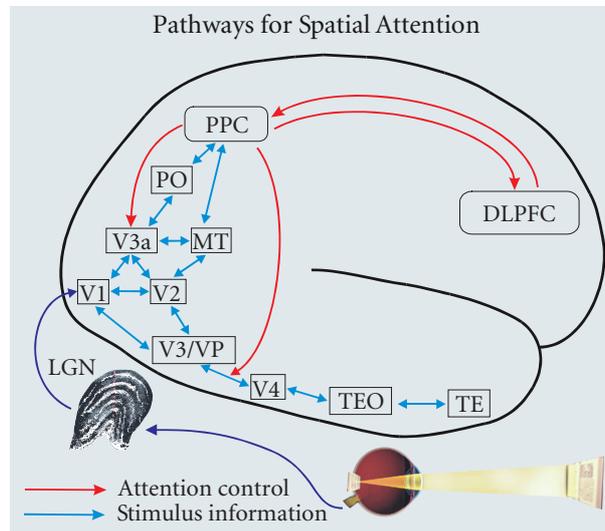


Fig. 19.1 Proposed pathways mediating spatial attention based on event-related potential (ERP), magnetoencephalography (MEG), and functional magnetic resonance imaging (fMRI) evidence. An attentional control network, consisting of interconnected dorso-lateral prefrontal (DLPFC) and posterior parietal (PPC) cortical areas, is proposed to modulate incoming visual information in both dorsal and ventral extrastriate areas. Evidence summarized by Martinez *et al.* (2001) suggests that attended inputs are enhanced dorsally in the region of areas V3/V3A/middle occipital gyrus and ventrally in the vicinity of area V4/fusiform gyrus in the time range 80–130 ms after stimulus onset. These enhanced signals are then fed forward to higher visual areas and back to lower areas, including V1.

network (Chapter 14, this volume). Neurophysiological studies in monkeys have documented strong attentional influences over visual processing in multiple extrastriate cortical areas (reviewed in Maunsell and McAdams 2000; Chapter 18, this volume), and recent experiments have found that neural activity in primary visual cortex (area V1) may also be modulated by attention under certain conditions (reviewed in Martinez *et al.* 2001). Neuroimaging studies have demonstrated that V1 activity may also be affected by spatial attention in humans; directing attention to a stimulus was found to increase neural activity at the retinotopic locus in area V1 corresponding to its location in the visual field (e.g. Tootell *et al.* 1998; Martinez *et al.* 2001). Kastner (Chapter 15, this volume) has presented new fMRI evidence that spatial attention can even influence neural activity at subcortical levels of the visual pathway. An increased BOLD signal was observed in the lateral geniculate nucleus (LGN) when subjects directed attention to a peripheral stimulus in comparison with counting letters at fixation. Kastner suggested that the LGN may serve as an ‘early gatekeeper’ that exerts a gain control over thalamo-cortical transmission of visual information.

In light of this evidence that spatial attention can influence neural activity in both the LGN and V1, one may well ask why Fig. 19.1 shows attentional control being exerted directly on higher, extrastriate levels of the visual–cortical pathway. The rationale for this comes from studies of event-related potentials (ERPs) and magnetoencephalography (MEG), which suggest that incoming visual information is first modulated at extrastriate levels rather than in V1 or the LGN. This apparent discrepancy between the results of ERP/MEG studies, on the one hand, and fMRI studies, on the other, may be resolved by considering that ERP/MEG recordings can determine the time course of attention-related neural activity with a millisecond-level of resolution, whereas blocked fMRI designs reveal neural activity patterns with high anatomical precision but give little information about when this activity occurs with respect to stimulus events.

19.3 Evidence from ERP/MEG experiments

Several ERP and MEG studies have investigated the effects of spatial attention on different components of the stimulus-evoked waveforms that have been ascribed to striate and extrastriate neural generators (reviewed in Martinez *et al.* 2001; Noesselt *et al.* 2002). In a recent study of this type (Di Russo *et al.* 2003), small, circular stimuli were flashed to right and left visual field positions in random order, either in the upper or lower quadrants (Fig. 19.2). The subject's task was to pay attention to the flashes on one side (ignoring the contralateral flashes) and to detect occasional target flashes of smaller diameter. The earliest ERP component (C1, onsetting at 50–60 ms after stimulus onset) was maximal over the midline parietal-occipital scalp (Fig. 19.2A) and was localized through dipole modeling to midline occipital cortex, which includes area V1 within the calcarine fissure (Fig. 19.2B, C). Significantly, the negative-going C1 was not affected by spatial attention. The earliest attentional modulation was an enhanced positivity (the P1 component onsetting at 70–80 ms) to attended stimuli that was largest over the lateral occipital scalp and was localized to generators in both dorsal and ventral extrastriate cortex (Martinez *et al.* 2001; Noesselt *et al.* 2002; Di Russo *et al.* 2003).

Several lines of evidence indicate that the attention-invariant C1 component does, in fact, represent early evoked neural activity in area V1. These include its being the earliest component in the visual ERP, its co-localization with fMRI activation in area V1 and, most importantly, its polarity inversion when comparing the ERPs to upper versus lower visual field quadrants. (Di Russo *et al.* 2003; Martinez *et al.* 2001). This inversion may be ascribed to the anatomical projections of the upper and lower visual fields to opposing cortical surfaces in area V1 within the calcarine fissure.

It remains to be explained why the C1 component that appears to represent early evoked activity in area V1 shows no modulation with attention, while the co-localized hemodynamic response seen with fMRI does indicate an attention-related increase in

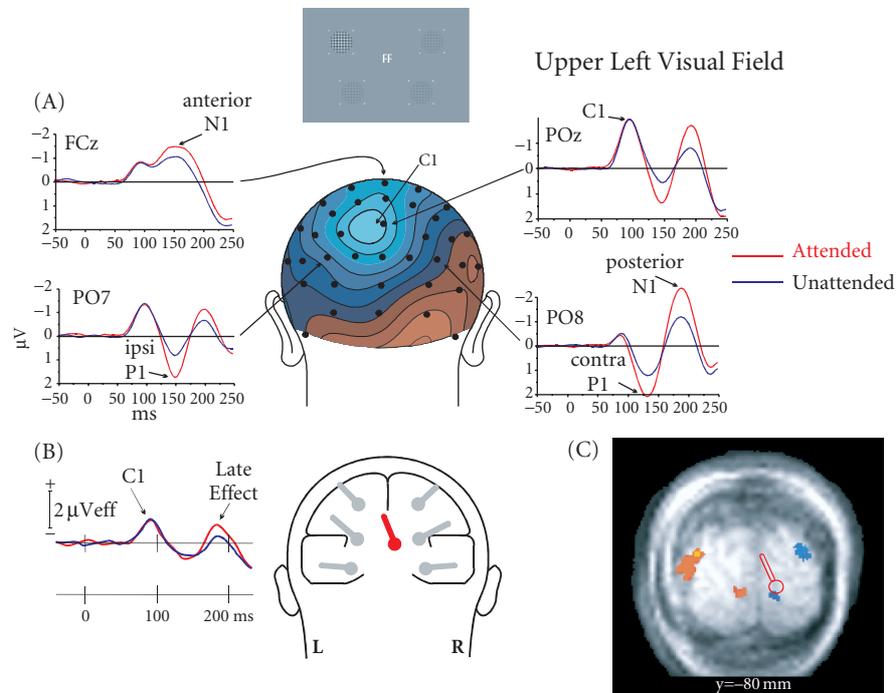


Fig. 19.2 (A) Averaged ERPs from four scalp sites in response to stimuli in the upper left quadrant in a spatial attention experiment by Di Russo *et al.* (2003). Superimposed ERP waveforms compare conditions when those stimuli were attended versus when stimuli in the opposite field were attended. Head map shows scalp distribution of the early negative C1 component, which is unaffected by attention (bluish shades are negative, reddish are positive). Note that P1 (80–130 ms) and N1 (130–200 ms) components are enhanced by attention. Averaged data from 23 subjects. (B) Dipole model of neural source of the C1 component in midline occipital cortex. Waveforms at the left show the time course of dipole source activity. Note enhanced source activity to attended stimulus at 160–220 ms (late effect). (C) C1 dipole is superimposed on a structural MRI averaged over seven subjects who performed the same task while undergoing fMRI. Blue/red pixels are those showing greater BOLD signals during attention to the left/right visual quadrants. Left hemisphere is on the left. Note spatial correspondence between the C1 dipole position and that of the attention-related BOLD signal.

neural activity in V1 (Martinez *et al.* 2001; Noesselt *et al.* 2002) (see Fig. 19.2C). Several possible explanations may be entertained to account for this dissociation:

- (1) the attention-related increase in V1 activity is the result of delayed feedback from higher extrastriate areas;
- (2) this activity represents an increase in the baseline firing rate of V1 neurons rather than an increase in stimulus-evoked activity (see Chapter 15, this volume);

- (3) the V1 activity occurs in oscillatory bursts that are not time locked to the attended stimuli and hence do not appear in averaged ERP waveforms; or
- (4) the V1 activity takes place in stellate neurons that do not produce far-field electrical or magnetic signals.

Recent studies in monkeys provide evidence for the first (delayed feedback) alternative. Single-unit recordings from V1 have shown that spatial attention typically enhances neural activity at fairly long latencies (80–100 ms or greater), well beyond the peak of the initial sensory evoked response (Vidyasagar 1999; Lamme *et al.* 2000). In addition, concurrent ERP recordings from implanted electrodes in multiple visual areas showed that evoked activity in higher extrastriate areas such as V4 was enhanced by attention at shorter latencies than was activity in area V1 (Schroeder *et al.* 2001).

ERP and MEG recordings in humans are also supportive of a delayed feedback to area V1 following attended stimuli. Dipole modeling of ERP and MEG attention effects have found that a source localized to area V1 shows a late (onset at 130–160 ms) but not an early modulation with attention (Martinez *et al.* 2001; Noesselt *et al.* 2002). This effect can be seen in Fig. 19.2B, where the source waveform of the calcarine dipole shows a delayed modulation with attention. Taken together, these animal and human studies suggest that spatial attention first enhances sensory inputs from attended locations in higher extrastriate areas, and these augmented signals are then conveyed back to area V1 by re-entrant feedback projections (see Fig. 19.1). Such feedback signals are proposed to improve figure/ground segregation and the salience of stimuli within the spotlight of spatial attention (Lamme *et al.* 2000).

As for the reported increase in LGN activity with spatial attention (Chapter 15, this volume), it would seem that further experiments are required to determine whether this activity reflects increased feedforward signals, delayed feedback signals, or a shift in baseline activity. The finding from ERP studies that the initial C1 is not affected by attention appears to weigh against a feedforward gain control at the level of the LGN, but differences in experimental design between the ERP and fMRI studies make it difficult to rule out this possibility altogether.

19.4 Mechanisms of spatial selection

Reynolds and Fallah (Chapter 18, this volume) describe how the biased-competition model introduced by Desimone and Duncan (1995) can account for a good deal of the attention effects observed in monkey single-unit studies. In particular, when multiple stimuli are presented in close spatial proximity so as to fall within a neuron's receptive field, attention can bias the competitive interactions between the stimuli so that one input exerts a dominant influence over the cell's firing. Kastner (Chapter 15, this volume) has reported analogous competitive interactions between neighboring stimuli using fMRI, and by varying their spatial separation it was possible to estimate the dimensions of receptive fields in different visual areas of humans.

The biased-competition model seems to provide a good account of attentional selections among neighboring stimuli, but its application is less obvious when attended stimuli and distractors are widely separated and hence minimally competitive. How, then, can we explain the substantial modulations of neural activity that have been observed in ERP, fMRI and single-unit studies under conditions where attended and ignored stimuli were presented in opposite visual fields and separated by 8–10° or more (e.g. Hillyard and Anllo-Vento 1998; Tootell *et al.* 1998; Maunsell and McAdams 2000). Kastner (Chapter 15, this volume) has proposed a mechanism for attentional selection under such conditions, which involves a ‘push–pull’ facilitation of inputs from the attended location together with a suppression of inputs from outside the attended location. This suppressive effect was postulated to decrease systematically as a function of distance from the locus of attention and to be mediated by long-range horizontal and transcallosal connections.

While Kastner’s push–pull hypothesis appears to have considerable explanatory power, further work is needed to clarify its neural underpinnings. In particular, the role of the frontal–parietal control network in allocating attention across the visual field needs to be explicated. It may well be the case that top-down projections from this network are primarily responsible for establishing the zones of excitatory and inhibitory bias over widely separated regions of the visual field. The studies reviewed by Driver *et al.* (Chapter 14, this volume) indicate further that spatial attention has an intrinsically multimodal organization, such that directing attention to a stimulus in one modality facilitates the processing of stimuli in multiple modalities at the attended location. This suggests that the top-down control network imposes concurrent patterns of facilitatory and inhibitory influences within the sensory cortices of multiple modalities and that these patterns are kept in spatial register with one another.

19.5 Object-based selection

As noted by several contributors (Chapters 14, 16, and 18, this volume), it is now evident that attention may be directed towards, and guided by, objects as well as locations in the visual field. For example, it has been found that attention can be divided more efficiently between different aspects of the same object than between separate objects, and that attention directed towards one part of an object may extend automatically to its other parts. In a series of ingenious experiments, Valdes-Sosa and co-workers (Chapter 16, this volume) have investigated the neural mechanisms underlying object-based selection in humans by means of combined psychophysical/ERP experiments. A key feature of their designs has been to trigger ERPs by transforming the shape of the object being attended, rather than by superimposing a transient stimulus upon it that may not be integrated into the object’s perceptual representation. In one such experiment it was found that a shape change was judged more accurately when the shape belonged to the same object in which a preceding shape change had occurred, as opposed to a different object.

Recordings of ERPs in this experiment revealed that the early N1 component (at a latency of around 200 ms) was reduced when successive shape changes occurred within different objects (relative to the same object). The scalp topography of this N1 resembled that of the N1 previously found to be modulated by purely spatial attention, which appears to be generated in extrastriate visual areas (Clark and Hillyard 1996; Martinez *et al.* 2001; DiRusso *et al.* 2003). Accordingly, Valdes-Sosa *et al.* suggested that the different-object cost may be associated with an early suppression of sensory information in extrastriate cortex. An important question for future study is whether this object-based selection is achieved by the same mechanisms that enable purely spatial selection—that is, does spatial attention automatically spread from an attended part of an object throughout the entire object? This could be investigated by recording ERPs to shape-change stimuli when attention is explicitly directed to their locations, and comparing the scalp distributions and neural sources of the ERPs modulated by space-based and object-based attention.

In other experiments Valdes-Sosa and colleagues (Chapter 16, this volume) have shown that attention can be directed selectively to one of two spatially superimposed moving dot patterns having the appearance of overlapped transparent surfaces. Remarkably, although the design precludes selection by location, the ERP triggered by displacement of the attended surface showed enlarged P1/N1 components that resembled the ERP modulations reported to occur during spatial attention. Again, it would be of great interest to compare the properties of these ERP modulations under conditions of spatial versus surface-based selection, to determine whether common mechanisms are involved. Further insight into the cortical mechanisms of surface-based attention comes from preliminary studies in monkeys by Reynolds and Fallah (Chapter 18, this volume), who found that neurons in area V4 can be driven selectively by a transparent surface that captures attention. Taken together, these human and animal studies suggest that attentional selections based on locations, features, and objects may all be carried out in low-level extrastriate visual areas, presumably under the influence of task-specific top-down and feedback influences from higher cortical areas.

19.6 Conclusion

From this brief overview, it is apparent that different imaging methods must be combined in common experiments if we are to achieve a full understanding of where, when, and how sensory information is modulated by the brain's attentional systems. While hemodynamic imaging with fMRI can give a detailed anatomical picture of the brain regions participating in attentional selection, methods with better temporal resolution, such as ERP/MEG, are needed to reveal the timing and sequencing of neural events in the different regions. In particular, we cannot assume that neural activity at an anatomically early stage of the sensory pathways necessarily occurs earlier in time than activity at a higher level in the cortical processing hierarchy. There is mounting

evidence that feedback from higher to lower anatomical levels plays an essential role in visual perception and attention and acts in concert with feedforward signals to form a bidirectional cascade of sensory information.

References

- Clark, V.P. and Hillyard, S.A. (1996). Spatial selective attention affects early extra striate but not striate components of the visual evoked potential. *Journal of Cognitive Neuroscience*, **8**, 387–402.
- Desimone, R. and Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, **18**, 193–222.
- Di Russo, F., Martinez, A., and Hillyard, S.A. (2003). Source analysis of event-related cortical activity during visuo-spatial attention. *Cerebral Cortex*, **13**, 486–99.
- Hillyard, S.A. and Anllo-Vento, L. (1998). Event-related brain potentials in the study of visual selective attention. *Proceedings of the National Academy of Sciences*, **95**, 781–87.
- Lamme, V.A.F., Super, H., Landman, R., Roelfsema, P.T., and Spekreijse, H. (2000). The role of primary visual cortex (V1) in visual awareness. *Vision Research*, **40**, 1507–21.
- Martinez, A., Di Russo, F., Anllo-Vento, L., Sereno, M.I., Buxton, R.B., and Hillyard, S.A. (2001). Putting spatial attention on the map: timing and localization of stimulus selection processes in striate and extrastriate visual areas. *Vision Research*, **41**, 1437–57.
- Maunsell, J.H.R. and McAdams, C.J. (2000). Effects of attention on neuronal response properties in visual cerebral cortex. In M. Gazzaniga (Ed.), *The new cognitive neurosciences*. MIT Press, Cambridge MA.
- Noesselt, T., Hillyard, S.A., Woldorff, M.G., Schoenfeld, A., Hagner, T., Jancke, L., Tempelmann, C., Hinrichs, H., and Heinze, H.J. (2002). Delayed striate cortical activation during spatial attention. *Neuron*, **35**, 575–87.
- Schroeder, C.E., Mehta, A.D., and Foxe, J.J. (2001). Determinants and mechanisms of attentional modulation of neural processing. *Frontiers in Bioscience*, **6**, 672–84.
- Tootell, R.B.H., Hadjikhani, N., Hall, E.K., Marrett, S., Vanduffel, W., Vaughan, J.T., and Dale, A.M. (1998). The retinotopy of visual spatial attention. *Neuron*, **21**, 1409–22.
- Vidyasagar, T.R. (1999). A neuronal model of attentional spotlight: parietal guiding the temporal. *Brain Research Reviews*, **30**, 66–76.

Tactic: Add Visual Distinctions to High ROI Products. Attention is correlated with choice. People are more likely to choose an option if they spend more time looking at it (Atalay, Bodur, & Rasolofarison, 2012). But you can also capture attention through static images of people. Those images are social cues – thus activating our STS region (Allison, Puce, & McCarthy, 2000). Here are the most important features Visual attention models have been used in many computer vision applications (Pal, 2016), including image and video compression and retrieval (Ouerhani et al., 2001; Li et al., 2011), multimedia technologies (Le Callet and Niebur, 2013), and 2D (Akamine and Farias, 2014) and 3D (Jiang et al., 2014) image quality assessment. In all examples the visual search for the target starts with the green circle which indicates the first hit, and continues up to six points on the image creating a scan path which simulates saccades.

FIGURE 1. Computer Science > Machine Learning. Title: Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. Authors: Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhutdinov, Richard Zemel, Yoshua Bengio. (Submitted on 10 Feb 2015 (v1), last revised 19 Apr 2016 (this version, v3)). Abstract: Inspired by recent work in machine translation and object detection, we introduce an attention based model that automatically learns to describe the content of images. We describe how we can train this model in a deterministic manner using standard backpropagation