



Cortical mechanisms of visual context processing in singleton search

Anna Schubö^{a,b,*,1}, Elkan G. Akyürek^{a,c,1}, En-Ju Lin^a, Ignacio Vallines^a

^a Ludwig Maximilian University, Munich, Germany

^b Philipps University, Marburg, Germany

^c University of Groningen, Groningen, The Netherlands

ARTICLE INFO

Article history:

Received 28 April 2011

Received in revised form 18 June 2011

Accepted 12 July 2011

Keywords:

Functional imaging

Visual attention

Texture segmentation

Context homogeneity

Visual search

ABSTRACT

When searching for a target object presented in a context of other, irrelevant objects, the dissimilarity between target and surrounding context elements as well as the similarity between context elements themselves affect search efficiency. The present functional imaging study explored the cortical mechanisms involved in processing the same target when surrounded by context arrangements of varying homogeneity. Results showed that brain activity increased in the precuneus, cingulate gyrus, and the middle temporal gyrus as context homogeneity and local feature contrast increased. Contexts with low homogeneity and local feature contrast, compared to contexts with high homogeneity and local feature contrast, increasingly involved areas near the corpus callosum and the medial frontal gyrus. The results support the assumption that contextual grouping and local target detection both contribute to perform the visual search task.

© 2011 Elsevier Ireland Ltd. All rights reserved.

The efficiency of searching for a specific, relevant target object among other, irrelevant objects depends to a large extent on the physical distinctiveness of the target from the irrelevant context elements. The more distinct the relevant target is from the context, the faster will the observer detect it [23]. The physical characteristics of the context elements themselves also matter: More homogeneous context elements make the target stand out from the background. Duncan and Humphreys [8] first emphasized the role of context elements in target detection. According to their attentional engagement theory, search efficiency is influenced by the similarity between target and context elements and the similarity among the context elements [8]. The similarity between target and context elements *decreases* search efficiency because context elements may, as they overlap with the target features, receive similar attentional “weights” as the target. The similarity of context elements, on the other hand, *increases* search efficiency as similar context elements can be grouped and then rejected faster as respective non-target elements. Several studies have shown that detecting the same target in various contexts depends on the target-context distinctiveness and on the similarity among the context elements [8,20].

The present study investigated the effects of context homogeneity on target-context segmentation by using functional magnetic resonance imaging (fMRI). In visual areas, segmentation may result from the integration of horizontal communication across neurons or from feedback projection from higher-order brain regions. In these regions, segmentation could result from neurons with larger receptive fields via the integration of information from lower-order areas. Higher-order region involvement in segmentation was expected because of relatively late modulations (>200 ms) previously observed in the EEG [20]. Further support was provided by Hayakawa and colleagues [11], who combined functional MRI and MEG to investigate the neural bases of target-context segmentation in an orientation singleton search task. Neuronal activity around the calcarine sulcus (CaS) was observed, which consisted of two subsequent peaks with the later peak (215 ms) considered to reflect target-context segmentation. This second activation peak was observed in roughly the same time window as posterior N2 effects in an ERP experiment by Schubö et al. [20] that were assumed to mirror efficient distractor rejection.

Functional imaging studies on target-context segmentation furthermore provided evidence for higher-order extrastriate visual area involvement [e.g., 2,3,13,15,16,22]. Kastner and colleagues [13] identified correlates of texture segregation in areas V4 and TEO but not in V1 (see also [19,22]). Hopf and coworkers [12] found a significantly smaller BOLD response for salient pop-outs compared to non-salient stimuli in V1 but not in higher-order visual areas, but only when the stimuli were attended. Leonards and colleagues [17] investigated the cortical networks involved in simple, parallel vs. complex, serial visual search and observed activity in

* Corresponding author at: Experimental and Biological Psychology, Philipps University Marburg, Gutenbergstraße 18, D-35032 Marburg, Germany.

Tel.: +49 06421 28 25594; fax: +49 06421 28 28948.

E-mail address: anna.schuboe@staff.uni-marburg.de (A. Schubö).

¹ Shared first authorship.

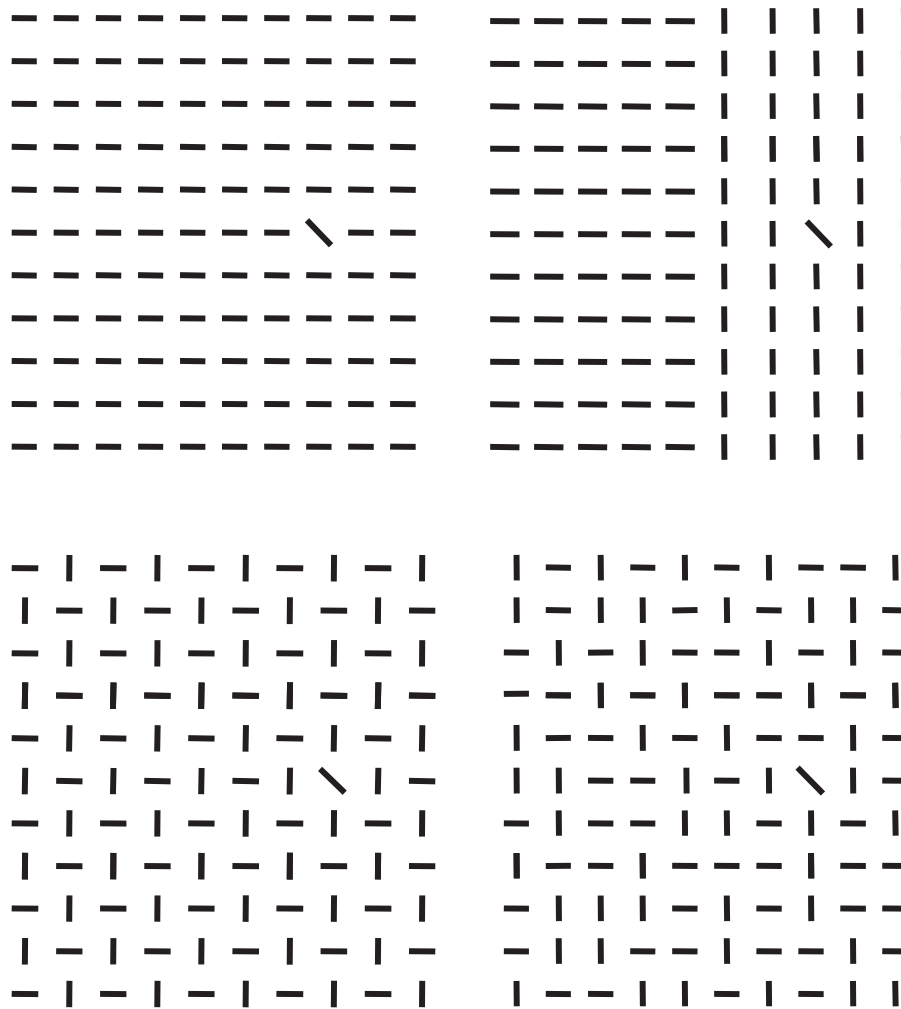


Fig. 1. Example arrays of homogeneous contexts (H, upper left), heterogeneous grouped contexts (G, upper right), heterogeneous pattern contexts (P, lower left), and heterogeneous random contexts (R, lower right).

extrastriate regions when participants performed pop-out search. Larsson and coworkers [16] used orientation adaptation as a tool to localize populations of neurons selective for orientation contrasts. They found that orientation-selective response adaptation was significantly stronger in extrastriate visual areas than in area V1 and concluded that stimulus orientation contrast was extracted by additional processing after V1. Finally, Altmann and colleagues [2] found that neurons in the LOC encoded context information, but only when target-context segmentation was rather difficult.

To date, there has been no attempt to investigate the neural correlates of varying context homogeneity during target-context segmentation; an issue the present study sought to address. Participants were asked to detect target objects in surrounding contexts that consisted of the same type of elements but were arranged to produce contexts of various heterogeneity. Context elements could consist of the same type of elements only (i.e., either vertical or horizontal lines, varied between trials) forming a homogeneous context (Fig. 1, upper left) or they could consist of both types of elements (both vertical and horizontal lines presented simultaneously) that were either spatially grouped (forming a heterogeneous yet grouped context; Fig. 1, upper right) or alternating (forming a heterogeneous context with a regular pattern; Fig. 1, lower left) or intermixed randomly in the display (forming a heterogeneous context with a random element arrangement; Fig. 1, lower right). Since the classification of line elements remained the same for all four contexts (oblique lines were always targets, and vertical and

horizontal lines always context elements), neither target identification nor the target-distractor similarity relation differed among these conditions. However, different grouping processes may have occurred and affected search efficiency. Increasing context heterogeneity decreases distractor similarity and thereby may reduce the possibility of similarity-based context grouping. Thus one would expect the difficulty of target detection in target-present trials and that of distractor rejection in target-absent trials to increase as background homogeneity decreases. Physiologically, this could be reflected in increased involvement of higher-order brain regions in more heterogeneous contexts. Second, an increase in context heterogeneity also increases the heterogeneity at the target location, i.e., the local feature contrast. As high local feature contrast is known to attract the observer's attention spontaneously [9,21], higher local feature contrast may support fast and spontaneous target detection. Thus, the difficulty of target detection should increase as local feature contrast decreases, which might be correlated with corresponding changes in brain regions involved in selection or spatial attention. Former fMRI studies have not systematically investigated the effects of global context homogeneity and local feature contrast on target detection. The present study aimed to investigate their respective contributions to the observed brain activity.

Twelve paid volunteers (mean age 22.7 years) participated in the experiment. All had normal or corrected-to-normal vision and reported to have no history of neurological problems. Participants

performed a 30-min pretest outside the scanner to familiarize themselves with the task. The experimental protocol was designed and implemented in accordance with the Declaration of Helsinki. All participants gave written informed consent.

A 3 T Siemens Allegra head MR scanner equipped with a gradient strength of 40 mT/m and ultrafast slew rate of 400 T/m/s was used to acquire both high-resolution T1-weighted anatomical images and T2*-weighted echo planar images (EPI) sensitive to blood oxygenation level-dependent (BOLD) contrast during the experiment. The stimuli were projected by a JVC DLA-G20 digital projector on a translucent screen mounted inside the scanner gantry and could be seen by subjects through a set of mirrors mounted on the head coil. The distance from the screen to the participants' eyes was about 60 cm and the radius of the screen was 34 cm. A response box was placed in the participants' dominant hand so that they could respond comfortably with their index and middle fingers.

Stimulus arrays contained 110 black lines (1° of visual angle in length and approximately 0.05° in width) arranged on an imaginary matrix consisting of 10 columns \times 11 rows of evenly spaced cells. Stimuli were shown in black on a light-gray background. The context consisted of vertical and horizontal line elements that were distributed differently in the four context conditions (Fig. 1). Homogeneous contexts (denoted with "H") consisted of either horizontal lines or vertical lines only, presented with equal probabilities across trials. Heterogeneous contexts consisted of both horizontal and vertical lines, arranged either separately in the two visual hemifields (grouped contexts; "G") or alternating across the visual field (pattern contexts; "P") or randomly arranged (without any constraints, randomly assigned to each element position; "R") across the visual field (random contexts). In target-present trials, a single oblique line element tilted 45° to the left (the target) was presented at the middle horizontal line of the matrix 3.9° to the left or right from the centre of the screen. A mask, presented immediately after the search array, was constructed by superimposing horizontal, vertical, left- and right-tilted line elements at each matrix cell.

Each trial started with the presentation of a central fixation cross that remained on the screen for 300 ms followed by the search array presented for 100 ms and subsequently masked. After 600 ms, the mask disappeared and the fixation cross of the next trial was presented. Participants were asked to press the right response key if they had detected a target and the left key if they had not. Target-absent and target-present trials were presented with equal probability. Participants were asked to respond as accurate and as fast as possible. Each block consisted of 32 trials, and one complement of all four context conditions formed one experimental cycle. Between blocks, an 18-s rest period was given (720 s total). In each cycle, the four context conditions were mixed randomly. A scanning session contained ten cycles for each participant (1280 s total). Individual trials entered behavioral analysis, and blocks were used for physiological analysis.

T2*-weighted echo-planar images with 30 transversal slices were acquired (interleaved) in each scan (field of view: 192×192 mm, matrix resolution: 64×64 , TE: 30 ms, TR: 2 s, flip angle: 85°) in order to cover the largest volume of the cerebrum. Each of the 30 contiguous slices consisted of 64×64 three mm isotropic voxels. One thousand continuous measurements (time frames) were acquired for each participant. During functional data scanning, behavioral performance (errors and response times) was recorded. After functional scanning, a whole-brain T1-weighted anatomical image, which contained 160 transversal slices with a voxel size of $1 \times 1 \times 1$ mm, was acquired for each subject.

SPM5 (Wellcome Trust Centre for Neuroimaging, London UK) was used for functional and anatomical data preprocessing and for statistical analysis. T2* functional images were first motion-corrected to the first volume, time-corrected using the middle slice as a reference, and then co-registered to the T1 anatomical image.

Motion over the course of the scanning session was typically below 2 mm in either plane, and rotation was below 2° . Normalization (using a standard EPI template volume based on the MNI reference brain; [7]) and spatial smoothing (with an isotropic 6 mm FWHM Gaussian kernel) were performed afterwards. Voxels that were activated during the experimental conditions were identified by statistically testing BOLD signal levels against a model constructed by convolving box-car function with a canonical hemodynamic function that modeled the accumulated responses produced by trials in each block of every condition. High-pass filtering removed participant-specific low-frequency drifts in signal, and global changes in activity were removed by using proportional scaling. Linear contrasts between the different regressors representing the experimental conditions allowed determination of activated areas by creating a spatially distributed map of the t statistic. Comparisons were performed between the four context conditions in order to create statistical maps of each contrast for each subject. With subjects as the random factor, the contrast maps were further analyzed for each comparison separately in order to determine the regions of activation. The significance level was set at $p < .001$ (uncorrected) and the extent threshold was set at a cluster size of $k > 10$.

Percentage of correct responses and response times (RT) were registered for target-present and -absent trials and the four context heterogeneity conditions. Trials with RTs shorter than 200 ms and longer than 800 ms were excluded from RT analysis. Repeated-measures ANOVAs were conducted separately for (correct) RTs and percentage of correct responses on the factors TRIAL TYPE (target-present vs. target-absent) and CONTEXT (homogeneous vs. grouped vs. pattern vs. random).

The percentage of correct responses was higher in target-absent ($91.5 \pm 2.2\%$) than in target-present trials ($84.6 \pm 2.7\%$), $F(1, 11) = 11.97$, $p < .01$, $\eta_p^2 = 0.5$. The omnibus ANOVA showed no overall difference for context heterogeneity ($p > .15$), but repeated contrasts revealed a significant decrease in performance between pattern ($87.7 \pm 2.9\%$) and random contexts ($85.8 \pm 3.1\%$), $F(1, 11) = 5.36$, $p < .05$, $\eta_p^2 = 0.33$ (homogeneous: $88.8 \pm 1.9\%$; grouped: $90.0 \pm 1.7\%$).

Response times differed for homogeneous (408 ± 13.7 ms), grouped (411 ± 12.3 ms), pattern (433 ± 12.1 ms) and random contexts (431 ± 14.0 ms), $F(3, 33) = 28.33$, $p < .001$, $\eta_p^2 = 0.72$. Repeated contrasts revealed that there was a significant response time increase between grouped and pattern contexts (21.7 ms), $F(1, 11) = 34.41$, $p < .001$, $\eta_p^2 = 0.76$, and between grouped and random contexts (19.3 ms), $F(1, 11) = 26.95$, $p < .001$, $\eta_p^2 = 0.71$, other comparisons $p > .3$. Responses were overall slightly faster in target-absent (417 ± 14.5 ms) than in target-present trials (425 ± 12.1 ms), but this effect differed among the four context conditions, as indicated by a significant TRIAL TYPE \times CONTEXT interaction, $F(3, 33) = 8.77$, $p < .005$, $\eta_p^2 = 0.44$. Post hoc t -tests revealed faster target-absent RT compared to target-present RT in homogeneous contexts ($\Delta 20.4 \pm 6.4$ ms), $t(11) = 3.20$, $p < .01$, and in grouped contexts ($\Delta 16.5 \pm 5.8$ ms), $t(11) = 2.85$, $p < .05$, but not in pattern and random contexts ($p > .3$).

Bi-directional contrasts were first set up between homogeneous and grouped contexts, but this comparison did not reveal clusters that were differentially activated either way (H > G and H < G). The comparison between pattern and random contexts likewise did not reveal significant differences (both P > R and P < R). Thus, further tests were set up to directly compare these pairs; HG > PR, and HG < PR.

As shown in Fig. 2A (cluster 1), the cingulate gyrus in the right hemisphere showed clearly higher activity in H and G contexts than in P and R contexts ($t(11) = 7.33$, $p < .001$, MNI 2, -32, 30; $k = 79$; contrast estimate [CE] = .258). The cingulate gyrus has been shown to be involved in attentional target processing [24], which fits well with a

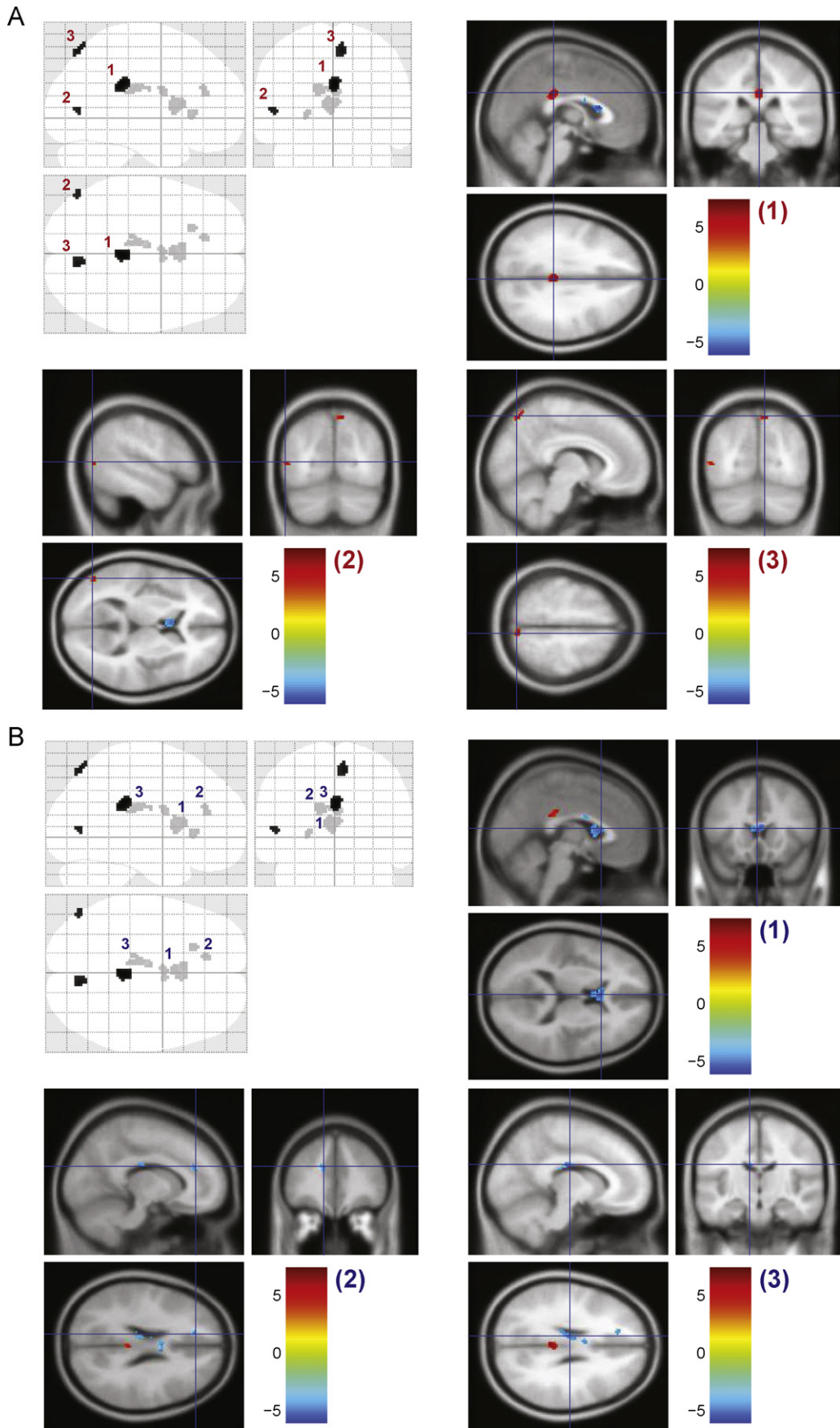


Fig. 2. (A) Statistical maps of the probability that areas were significantly more activated by homogeneous and grouped contexts than by pattern and random contexts. (B) The same for significantly higher activation induced by pattern and random contexts than by homogeneous and grouped contexts. Color intensity as shown on the legend represents the value of the *t* statistic.

facilitation of attentional deployment when context elements can be grouped and rejected efficiently. Smaller clusters of increased activity were located in the middle temporal gyrus (MTG) in the left hemisphere (cluster 2: BA 39; $t(11)=4.68$, $p<.001$, MNI -52 , -72 , 8 ; $k=13$; CE=.347), and in the precuneus in the right hemisphere (cluster 3: BA 7; $t(11)=5.65$, $p<.001$, MNI 8 , -72 , 58 ; $k=33$; CE=.378), an area associated with general attentional processing and the encoding of spatial locations [5,10].

As shown in Fig. 2B, P and R contexts elicited higher activity in the medial frontal gyrus (MFG) than H and G contexts (cluster 2: $t(11)=4.96$, $p<.001$, MNI -12 , 38 , 24 ; $k=21$; CE=.176). The MFG has been associated with the specification of the target template [1], which may have been more demanding in the pattern and random contexts. Increased activity in P and R contexts was also observed in four clusters in and nearby the corpus callosum and extra-nuclear areas: $t(11)=5.69$, $p<.001$, MNI -22 , 26 , 4 ; $k=20$; CE=.205; cluster 1: $t(11)=5.99$, $p<.001$, MNI -2 , 18 , 12 ; $k=133$; CE=.305; $t(11)=4.93$, $p<.001$, MNI 4 , 2 , 22 ; $k=26$; CE=.192; cluster 3: $t(11)=5.49$, $p<.001$, MNI -10 , -16 , 26 ; $k=43$; CE=.15. These clusters may have been involved due to increased inter-hemispheric communication with irregular contexts.

The present experiment investigated the effect of different context heterogeneity on the detection of an orientation singleton target. Both the global heterogeneity among context elements, and the local feature contrast around the target location was varied. It was assumed that both factors could affect search behavior and contribute to efficient orientation singleton detection.

Behavioral performance increased as context heterogeneity decreased, which was visible in RT and accuracy. A significant decrease in RTs was observed for homogeneous and grouped contexts compared to pattern and random contexts. This is in accordance with the assumption that local feature contrast results from the increase in similarity of context elements in the area around the target. Interestingly, this improvement was observed not only in target-present but also in target-absent trials. Thus local feature contrast may generate both the pop-out phenomenon in target present trials and speed performance in target absent trials without necessarily changing the search mode, as previous work has clearly shown that there is no evidence for serial search in even completely random contexts [20].

The fMRI data showed increased involvement of the cingulate gyrus, the precuneus, and the middle temporal gyrus when context heterogeneity was lower. Despite the apparent regularity in the arrangement of elements in the pattern contexts, there was no difference between pattern and random contexts, suggesting that the former were processed similarly to contexts with randomly arranged elements. Thus the regularity of the arrangement itself seems to be of minor relevance, and could not be used to segment the background as efficiently as the homogeneous and grouped contexts, which consisted of two element types that were homogeneously arranged in large parts of the visual field in a way that neighboring elements were always identical. The present results suggested that similarity of elements in texture-like stimuli is one precondition for successful grouping [19].

Thielscher and colleagues [22] suggested that midlevel visual areas form a key stage in figure-ground segregation. They assume that cells in mid-level areas such as area V4 with larger receptive fields play a central role in the robust identification of texture boundaries also under noisier background conditions: Texture borders are primarily detected in mid-level visual areas with LOC involvement; with the LOC reflecting a subsequent processing stage [2,15]. The present results showed increased activation of the middle temporal gyrus in the processing of more homogeneous contexts, which did not quite match the LOC coordinates found in the literature [e.g., 14]. This suggested that in our task the more advanced stage of processing that is executed in the LOC

was not necessary, probably due to the relative simplicity of our stimuli. There is some evidence that the differential activity found in the present task is related to early stages of object recognition (i.e., pre-LOC), rather than for instance spatial attention: The study by Arrington and colleagues [3] showed increased activity in the superior temporal gyrus for object-based cueing compared to location-based cueing in a comparatively simple visual search task.

Contexts in which local feature contrast at the target location was high furthermore led to a facilitation of attention allocation towards the target. Conversely, the increased activity elicited by more heterogeneous contexts in which local feature contrast was low seemed to indicate that increased inter-hemispheric communication was required, as evidenced by the increased involvement of neurons in the vicinity of the corpus callosum [4,6,18]. This might have been expected, given the variability between the left and right visual hemifields in these contexts; in the homogeneous and grouped contexts, the layout of the stimuli was more predictable (or identical) on both sides of the visual field. The involvement of the cluster in the medial frontal gyrus when background contexts were more heterogeneous might reflect increased demand on processes associated with maintaining the target template for successful selection [1]. In more homogeneous contexts target selection may have been easier because pop-out effects were facilitated. In such cases, it is not as essential to have a strong target template, since one can easily find the target by just looking for anything salient. In the more heterogeneous context, increased effort may thus have been needed to locate the target.

Homogeneity of the context elements leads to fast similarity grouping and subsequent rejection as target-absent on target-absent trials and to singling out the element that does not belong to the uniform structure in target present trials. Thus context grouping provides a strong tool for fast and efficient processing of large structural units. High local feature contrast may guide attention to the target location and thereby facilitate the target template matching process. Duncan and Humphreys [8] considered target-template matching and element grouping to be distinctive mechanisms. The present results are compatible with this assumption, as contextual grouping and target template matching elicited differential activity in different brain regions; the former tied to the processing of global shape, and the latter linked to attentional allocation to the target.

Acknowledgement

This work was supported by the German Research Foundation (DFG) as part of the Excellence Cluster “Cognition for Technical Systems” (CoTeSys), Project #148 to AS.

References

- [1] E.G. Akyürek, I. Vallines, E.-J. Lin, A. Schubö, Distraction and target selection in the brain: an fMRI study, *Neuropsychologia* 48 (2010) 3335–3342.
- [2] C.F. Altmann, A. Deubelius, Z. Kourtzi, Shape saliency modulates contextual processing in the human lateral occipital complex, *J. Cogn. Neurosci.* 16 (2004) 794–804.
- [3] C.M. Arrington, T.H. Carr, A.R. Mayer, S.M. Rao, Neural mechanisms of visual attention: object-based selection of a region in space, *J. Cogn. Neurosci.* 12S2 (2000) 106–117.
- [4] J.S. Bloom, G.W. Hynd, The role of the corpus callosum in interhemispheric transfer of information: excitation or inhibition? *Neuropsychol. Rev.* 15 (2005) 59–71.
- [5] R. Cabeza, F. Dolcos, S.E. Prince, H.J. Rice, D.H. Weissman, L. Nyberg, Attention-related activity during episodic memory retrieval: a cross-function fMRI study, *Neuropsychologia* 41 (2003) 390–399.
- [6] B.P. Choudhury, D. Whitteridge, M.E. Wilson, The function of the callosal connections of the visual cortex, *Exp. Physiol.* 50 (1965) 214–219.
- [7] C.A. Cocosco, V. Kollokian, R.K.S. Kwan, A.C. Evans, BrainWeb: online interface to a 3D MRI Simulated Brain Database, *Neuroimage* 5 (1997) S425.
- [8] J. Duncan, G.W. Humphreys, Visual search and stimulus similarity, *Psychol. Rev.* 96 (1989) 433–458.

- [9] J.H. Fecteau, D.P. Munoz, Saliency, relevance, and firing: a priority map for target selection, *Trends Cogn. Sci.* 10 (2006) 382–390.
- [10] L. Frings, K. Wagner, A. Quiske, R. Schwarzwald, J. Spreer, U. Halsband, A. Schulze-Bonhage, Precuneus is involved in allocentric spatial encoding and recognition, *Exp. Brain Res.* 173 (2006) 661–672.
- [11] T. Hayakawa, N. Fujimaki, T. Imaruoka, Temporal characteristics of neural activity related to target detection during visual search, *Neuroimage* 33 (2006) 296–306.
- [12] J.M. Hopf, T. Noesselt, C. Tempelmann, J. Braun, M.A. Schoenfeld, H.J. Heinze, Popout modulates focal attention in the primary visual cortex, *Neuroimage* 22 (2004) 574–582.
- [13] S. Kastner, W.P. De, L.G. Ungerleider, Texture segregation in the human visual cortex: a functional MRI study, *J. Neurophysiol.* 83 (2000) 2453–2457.
- [14] Z. Kourtzi, H.H. Bühlhof, M. Erb, W. Grodd, Object-selective responses in the human motion area MT/MST, *Nat. Neurosci.* 5 (2002) 17–18.
- [15] Z. Kourtzi, N. Kanwisher, The human lateral occipital complex represents perceived object shape, *Science* 24 (2001) 1506–1509.
- [16] J. Larsson, M.S. Landy, D.J. Heeger, Orientation-selective adaptation to first- and second-order patterns in human visual cortex, *J. Neurophysiol.* 95 (2006) 862–881.
- [17] U. Leonards, S. Sanaert, P. VanHecke, G.A. Orban, Attention mechanisms in visual search – an fMRI study, *J. Cogn. Neurosci.* 12S2 (2000) 61–75.
- [18] S.J. Luck, S.A. Hillyard, G.R. Mangun, M.S. Gazzaniga, Independent attentional scanning in the separated hemispheres of split-brain patients, *J. Cogn. Neurosci.* 6 (1994) 84–91.
- [19] A. Schubö, E. Schröger, C. Meinecke, Texture segmentation and visual search for pop-out targets. An ERP study, *Cogn. Brain Res.* 21 (2004) 317–334.
- [20] A. Schubö, A. Wykowska, H.J. Müller, Detecting pop-out targets in contexts of varying homogeneity: investigating homogeneity coding with event-related brain potentials (ERPs), *Brain Res.* 1138 (2007) 136–147.
- [21] J. Theeuwes, Top-down search strategies cannot override attentional capture, *Psychon. Bull. Rev.* 11 (2004) 65–70.
- [22] A. Thielscher, M. Kölle, H. Neumann, M. Spitzer, G. Grön, Texture segmentation in human perception: a combined modeling and fMRI study, *Neuroscience* 151 (2008) 730–736.
- [23] J.M. Wolfe, T.S. Horowitz, What attributes guide the deployment of visual attention and how do they do it? *Nat. Rev. Neurosci.* 5 (2004) 1–7.
- [24] H. Yamasaki, K.S. LaBar, G. McCarthy, Dissociable prefrontal brain systems for attention and emotion, *Proc. Natl. Acad. Sci. U.S.A.* 99 (2002) 11447–11451.