NeuroImage 146 (2017) 71–80

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Scene segmentation in early visual cortex during suppression of ventral stream regions

Pablo R. Grassi[a,b,c,*], Natalia Zaretskaya[a,b,c], Andreas Bartels[a,b,c,*]

[a] *Centre for Integrative Neuroscience, University of Tübingen, 72076, Tübingen, Germany*
[b] *Department of Psychology, University of Tübingen, 72076, Tübingen, Germany*
[c] *Max Planck Institute for Biological Cybernetics, 72076, Tübingen, Germany*

ARTICLE INFO

Keywords:
fMRI
V1
Feedback
Gestalt
Ventral stream
Bistable perception

ABSTRACT

A growing body of literature suggests that feedback modulation of early visual processing is ubiquitous and central to cortical computation. In particular stimuli with high-level content that invariably activate ventral object responsive regions have been shown to suppress early visual cortex. This suppression was typically interpreted in the framework of predictive coding and feedback from ventral regions. Here we examined early visual modulation during perception of a bistable Gestalt illusion that has previously been shown to be mediated by dorsal parietal cortex rather than by ventral regions that were not activated. The bistable dynamic stimulus consisted of moving dots that could either be perceived as corners of a large moving cube (global Gestalt) or as distributed sets of locally moving elements. We found that perceptual binding of local moving elements into an illusory Gestalt led to spatially segregated differential modulations in both, V1 and V2: representations of illusory lines and foreground were enhanced, while inducers and background were suppressed. Furthermore, correlation analyses suggest that distinct mechanisms govern fore- and background modulation. Our results demonstrate that motion-induced Gestalt perception differentially modulates early visual cortex in the absence of ventral stream activation.

1. Introduction

The organization of a visual scene into a set of coherent objects is an extraordinary feat of the visual system. It involves the binding of local visual features into unified structures, segregation of fore- and background, and filling in of missing surface- and contour information. Monkey electrophysiology studies on perceptual grouping have shown that neurons in early visual cortex respond selectively to border-ownership and illusory contours (Peterhans and von der Heydt, 1989) and that foreground surfaces get filled in (Poort et al., 2012; Roelfsema et al., 2007). In accord with this, human fMRI studies have consistently shown modulation of early visual cortex during object or Gestalt processing, suggesting Gestalt-related feedback from higher level regions (Kok and de Lange, 2014; Murray et al., 2002). These findings are compatible with the framework of predictive coding, in that high-level predictions interact with sensory signals in early cortex (Friston, 2005; Rao and Ballard, 1999).

However, potential sources for this feedback have not been conclusively established. All shape or Gestalt stimuli previously used to report early visual modulation also involved activation of ventral shape or object processing regions in the ventral lateral occipital cortex (LOC) compared to control stimuli (Fang et al., 2008; Hirsch et al., 1995; Mendola et al., 1999; Murray et al., 2002; Stanley et al., 2003; Strother et al., 2012). Simple “Kanizsa” shapes (Mendola et al., 1999; Stanley and Rubin, 2003) as well as moving diamonds, 3D shapes, or structure-from-motion stimuli have been shown to activate LOC compared to their control stimuli (De-Wit et al., 2012; Fang et al., 2008; Halgren et al., 2003; Murray et al., 2003, 2002), and electrophysiology has identified grouping-related responses in V4 that may account for so-called filling-in through feedback (Cox et al., 2013). To our knowledge no prior study examined early visual cortex modulations during Gestalt percepts that are not accompanied by ventral stream activation.

In a prior study we had examined a bistable motion illusion that induced mutually exclusive percepts of either a dynamic illusory Gestalt or unbound local moving elements (Grassi et al., 2016; Zaretskaya et al., 2013). This stimulus was accompanied by suppression of early visual cortex and of ventral regions, with parietal cortex being the sole activated region and also causally involved in Gestalt percept generation (Grassi et al., 2016; Zaretskaya et al., 2013).

In the present study we examined the precise nature of early visual cortex modulation, as, in contrast to all preceding studies, LOC and

* Correspondence to: Vision and Cognition Lab, Centre for Integrative Neuroscience, University of Tübingen, Otrfried-Müller-Str. 25, 72076, Tübingen, Germany.
E-mail addresses: Pablo.grassi@cin.uni-tuebingen.de (P.R. Grassi), andreas.bartels@tuebingen.mpg.de (A. Bartels).

http://dx.doi.org/10.1016/j.neuroimage.2016.11.024
Received 20 July 2016; Accepted 10 November 2016
Available online 12 November 2016
1053-8119/

other ventral stream regions were suppressed during Gestalt perception. We examined regions of interest corresponding to the physical inducers, the foreground, the illusory borders, and the background, in areas V1 and V2. Despite the ventral stream suppression, we found a specific modulation pattern in the early visual cortex: illusory contour regions and foreground were enhanced, while the inducers and background were suppressed. In addition, the degree of modulation in different sub-regions correlated across V1 and V2, but not with each other, suggesting that enhancement of foreground and suppression of background is governed by distinct processes.

2. Material and methods

2.1. Participants

Seventeen volunteers (21–29 years old, ten females, three left-handed, one author) participated in this study after signing an informed consent form. All had normal or corrected-to-normal vision and no history of neurological impairments. The study was conducted according to the Declaration of Helsinki and was approved by the ethics committee of the University Clinic Tübingen.

2.2. Imaging parameters

Data were acquired on a 3 T Siemens Prisma system with a 64-channel head coil (Siemens, Erlangen, Germany). Functional data were acquired using a slice-accelerated multiband gradient-echo echo planar imaging (EPI) sequence using T2* weighted blood oxygenation level dependent (BOLD) contrast (multiband acceleration factor: 4, GRAPPA acceleration factor: 2), with the following parameters: repetition time (TR)=870 ms, echo time (TE)=30 ms, flip angle=56°, FOV=192×224 mm and an isotropic voxel size of 2×2×2 mm. This high-resolution sequence allowed us to measure the spatial distribution of BOLD responses in early visual areas with a higher accuracy compared to standard EPI sequences. Whole-brain functional images consisted of 56 slices. The initial ten images of each experimental fMRI run were discarded for equilibration of T1 signal. A high-resolution T1-weighted anatomical scan was performed for each participant (ADNI, 192 slices, voxel size 1 mm^3, TR=2 s, TE=3.06 ms, FOV=232×256 mm). Data for the main experiment and for the structural scan were acquired in the same session. Data for retinotopic mapping were acquired on a separate session.

2.3. Retinotopic mapping

Using standard retinotopic protocols we delineated areas V1–V4 using phase-encoding (Engel et al., 1994; Sereno et al., 1995) in seven subjects. For the remaining ten subjects, V1 and V2 were defined using anatomical labels generated by FreeSurfer (Hinds et al., 2009) and standard anatomical criteria (see ROI definition below) (Wandell et al., 2007). The stimulus used for retinotopic mapping consisted of a wedge shaped checkerboard (100% contrast, 6 Hz contrast inversion flicker, check sizes increased logarithmically with eccentricity, wedge covered 90°) on a gray background rotating around the fixation dot at the center of the screen with a cycle duration of 55.7 s. Image acquisition and preprocessing was identical to the main experiment. Surface reconstruction (Dale et al., 1999) as well as visual area definition was done in FreeSurfer (Fischl et al. 1999; http://surfer.nmr.mgh.harvard.edu/).

2.4. Stimulus, task, and experimental procedure

The bistable motion stimulus was generated using MATLAB 2010a (MathWorks, Natick, MA) with the Psychophysics Toolbox 3 extensions (Brainard, 1997; Pelli, 1997) on a Windows computer and presented using a linearized projector (1024×768 resolution, 60 Hz). The visual display covered 22×16.3° visual degrees and was viewed at 82 cm distance. The stimulus consisted of four pairs of dots moving in-phase on circular paths (Anstis and Kim, 2011) (see Fig. 1A). This stimulus is perceptually bistable, and can be perceived as local motion of individual dots or as two illusory squares sliding over each other (i.e. global Gestalt motion). These two percepts alternate during continuous viewing.

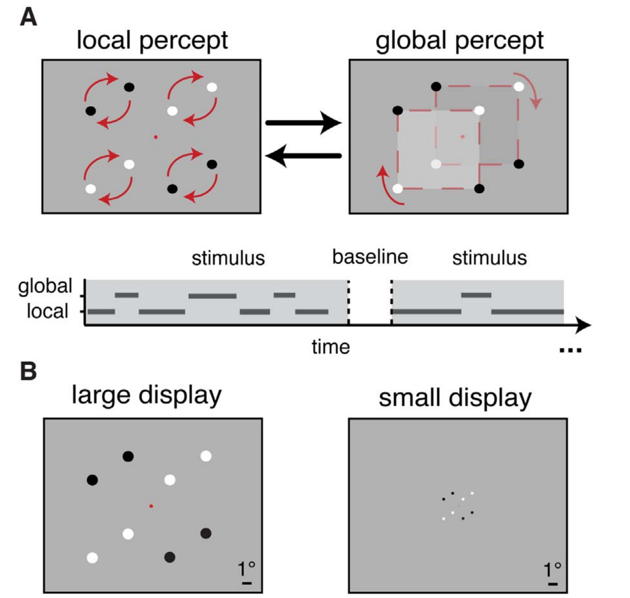

Fig. 1. Illustration of stimuli and the two possible perceptual interpretations. (A), The bistable motion stimulus led to alternations between the perception of local motion of dot pairs and of two illusory squares sliding over each other in planar motion (global Gestalt perception), respectively. During the stimulus presentation, subjects indicated periods of global or local motion perception by button presses. (B), Two stimulus conditions: large display and small display (ratio 1:4). Scale bar equals one visual degree.

During the experiment, subjects were asked to fixate the red dot and report their current percept by pressing and holding down one of two buttons on an fMRI-compatible keyboard with their right hand (one button was for the local, the other one for the global percept). Participants were instructed not to report anything if unsure of their percept. Simultaneous key presses were not used for further analysis.

The stimulus was presented in two different sizes in separate conditions (Fig. 1B). This allowed us to examine effects of eccentricity on BOLD responses. In the “large display” condition, individual dots had a size of one visual degree, the radius between dots and center of the dot pair was 2°, and the distance between each dot pair and center was 5°. The edge-length of the illusory square was 7°. Values in the “small display” were scaled to a fourth of those from the “large display” (i.e. 0.25° dot size, 0.5° dot-pair radius, 1.25° pair-center distance, edge-length 1.75°). Both stimuli had a red fixation dot (0.125°) in the center of the screen and were presented on a gray background (368.7 cd/m^2) at 100% contrast.

The dot rotation speed was adjusted for each participant individually prior to the experiment in order to achieve balanced durations of local and global perception. This led to a range of dot rotation speeds of 2–2.5 rotations per second (mean ± S.D.: 2.27 ± 0.13) across subjects. Dots from the same and opposite doublets had the same color (black or white) and alternated between each stimulus presentation.

Each fMRI run consisted of four stimulus presentations (90 s each) intermitted by 15 s of baseline consisting of fixation only. Every participant underwent 4–5 experimental runs. The experimental conditions (i.e. large and small stimulus) were presented in a pseudorandomized counterbalanced sequence.