Spatial Displacement, but not Temporal Asynchrony, Destroys Figural Binding

MANFRED FAHLE,* CHRISTOF KOCH†‡

Received 14 April 1994; in revised form 27 May 1994

What are the elementary features that the brain uses to bind spatially distinct parts in a visual scene into an unitary percept of an “object”? The Gestalt psychologists emphasized the extent to which motion, colour, luminance or spatial arrangement contribute towards object formation. Little is known about the role of time per se, rather than motion, in constituting an object. In particular, does the visibility or saliency of an object change if the various parts making up the object are not presented simultaneously? Using a simple experimental design, we show that very small spatial displacements can significantly influence the saliency of an object while large temporal asynchrony has no significant effect.

Time Figure–ground Rivalry Kanizsa triangle Synchronization

INTRODUCTION

The integration time of the human visual system is on the order of 100 msec (Barlow, 1958; Burr, 1980). Thus, images flashed at rates of 50 frames/sec or higher are perceived as stable, the basis for perceiving movies and television. Yet humans can reliably discriminate much smaller differences in temporal onset, an instance of temporal hyperacuity. Westheimer and McKee (1977) showed that subjects can reliably discriminate the order of onset of two small lines at the 3–5 msec level if the spatial separation between the two is in the range of 2–6 min arc. The ability of the human visual system to exploit temporal asynchronies for figure–ground segregation was studied by Fahle, Leonards and Singer (1993) and Fahle (1993), who presented subjects with a repeatedly flashed array of dots. If a rectangular set of dots is delayed by as little as 5–10 msec relative to the surrounding dots, the entire set of these points is perceived as “figure”. We conclude that humans clearly possess the ability to discriminate very small temporal delays, possibly involving the motion detection system (see also Ramachandran & Rogers-Ramachandran, 1991).

We wanted to study to what extent small temporal delays are used to discriminate among different percepts defined by spatially distinct features. Our motivation was trying to understand how the brain combines different aspects of objects, such as their position in space, their depth, colour, motion, the sounds emitted by them etc., into a coherently experienced “unitary percept”. This so-called binding problem (von der Malsburg, 1981; Crick & Koch, 1990) has received widespread attention with the discovery of spatially separate neurons in visual cortex of cats and monkeys that show oscillatory and temporally highly synchronized responses (Eckhorn, Bauer, Jordan, Brosch, Kruse, Munk & Reitböck, 1988; Gray, König, Engel & Singer, 1989; Kreiter & Singer, 1992).

In an abstract, Kiper, Gegenfurtner and Movshon (1991) report on using a threshold measurement to study the ability of the visual system to segment objects based on temporal cues. They find no difference when the elements of an array of oriented lines are presented synchronously or asynchronously with an array of differently oriented lines in the background. We here use a more sensitive test, exploiting the behaviour of bistable percepts. The idea of our experiment is simple. Certain visual stimuli, such as a Necker cube or many drawings of M. Escher, induce one of two possible percepts that alternate over time. Under symmetrical stimulus conditions both percepts are equally likely to be seen, but not at the same time. Fahle and Palm (1991) showed that of the two identical and partially overlapping Kanizsa triangles formed by illusory contours (Fig. 1), subjects almost always reported that one is “dominating” the other by partly occluding it. All of their subjects vividly experienced perceptual rivalry under dichoptic, monocular and binocular viewing conditions, with no significant bias towards seeing one or the other triangle. What would happen if we break the symmetry between the two figures making up the stimulus by varying the exact spatial arrangement of one triangle or by presenting the pacmen making up one triangle at different times? All things being equal, we expected any change that perturbs
the “saliency” of one triangle to result in a bias towards
the perception of the “perfect” triangle.

METHODS

We present nine subjects with one of three different,
randomly interleaved, stimulus configurations. Each
stimulus is presented repeatedly at a fixed frequency for
1 sec. In the standard stimulus (Fig. 1) the sides of the
equilateral triangle are 42 min arc long and the radius
of each pacman is 8.3 min arc, with an observation
distance of 2 m. All stimuli are displayed on an
oscilloscope screen controlled by two 16-bit D/A
converters driven at 1.1 MHz. Stimulus luminance is
140 cd/m² on a background of 15 cd/m², corresponding
to 80% contrast.

In the angle mode, we randomly increase the 60 deg
opening angle of one (or two) pacmen making up one
triangle by α and decrease the opening angle of the other
two (or one) pacmen by α. In the displacement mode,
we move the positions of the outer two pacmen of one
triangle by the same distance d in opposite directions. All
six pacmen in the previous two configurations are drawn
within less than 6 msec. In the temporal phase
configuration, the three pacmen constituting one triangle
are flashed simultaneously (“Triangle A”), followed by
the sequential presentation of pacmen 1, 2 and 3 of the
second triangle (“Triangle B”). The three intervals
between all four sets of (either one or three) pacmen are
constant (set to Φ). We randomly vary which of these four
sets is presented first (see also Fig. 2).

The angle, displacement and temporal phase con-
figurations are randomly interleaved and presented
repeatedly at frequencies between 5 and 75 Hz for a total
duration of 1 sec. These frequencies correspond to onset
asynchronies between 14 and 200 msec between sub-
sequent presentations of the stimuli. In a two-alternative
forced-choice task, the observers have to report which
triangle is the most “salient” or “dominant” one, without
being told which parameter is relevant. Subjects respond
“correctly” if they identify the triangle formed by the
three spatially perfect, simultaneously flashed pacmen.
Thresholds (here 75% correct responses) are determined
by the use of a staircase procedure [PEST (Taylor &
Creeelman, 1967)]. Each data point of each observer relies
on at least 100 presentations. Eight of the nine subjects are
naive regarding the aim of the study. None are provided
with any error feedback. Observers are asked to always
fixate a central point.

RESULTS

Breaking perceptual symmetry by changing the spatial
configuration of one triangle strongly enhances the
“saliency” of the unperturbed triangle [Fig. 2(a, b)].
Subjects usually report seeing the dominant triangle
partially occluding the other one. Averaged over all
observers and all stimulus presentation frequencies, the
threshold for spatial displacements is around 1.5 min arc
and the threshold for changes in the opening angle around
10 deg. That is, if the outer two pacmen of one triangle
each move by 1.5 min arc or the angle changes by more
than 10 deg, subjects perceive the “perfect” triangle as the
dominant one on at least 75% of the trials (notice that an
opening angle of 10 deg for a radius of 8.3 min arc
corresponds to an arc segment of 1.46 min arc). At fast
presentation frequencies (above 10 Hz), performance is
somewhat better than for low frequencies. We believe this
to be due to averaging taking place when the stimulus
is presented more than 10 times during the 1 sec
presentation time.

We could not measure any threshold in the temporal
phase configuration for any of the nine subjects at
frequencies above 5 Hz [Fig. 2(c)]. At these frequencies,
no subject ever identifies the synchronously flashed
stimulus with 75% probability or more. For frequencies
above 15 Hz, performance is not significantly different
from chance (at the 95% level; Student’s t-test). This is
true even if only the temporal phase configuration by itself
is tested rather than the mixed experiments with angular
and positional variation and the observers know the
discriminating feature (two observers). For 5–15 Hz
rates, individual performance scores are between 41% and
66%. At these slow frequencies and phase angles
of 90 deg, the individual flashed pacmen are 50 msec

FIGURE 1. Under symmetrical conditions, subjects report perceiving
one illusory Kanizsa triangle dominating the other one, usually
by partially occluding it. We present nine subjects with one of three
different, randomly interleaved, stimulus configurations. Each stimulus
is presented repeatedly at a fixed frequency for 1 sec (between 5 and
75 Hz for Fig. 2 and between 2.5 and 75 Hz for Fig. 3). In the angle mode,
we randomly increase the 60° opening angle of one (or two) pacmen
making up one triangle by α and decrease the opening angle of the other
two (or one) pacmen by α. In the displacement mode, we move the
positions of the outer two pacmen of one triangle by the same
distance d in opposite directions. All six pacmen in the previous two
configurations are drawn within less than 6 msec. In the temporal phase
configuration, the three pacmen constituting one triangle are flashed
simultaneously (“Triangle A”), followed by the sequential presentation
of pacmen 1, 2 and 3 of the second triangle (“Triangle B”). The three
intervals between all four sets of (either one or three) pacmen are
constant (set to Φ). We randomly vary which of these four sets is
presented first. Each data point of each of the nine observer relies on at
least 100 presentations.
(at 5 Hz) apart and can easily be seen sequentially, but subjects never report a sense of motion in these experiments. One subject was measured with a maximal phase angle of either 60° (AH) or 90° (FAH) with no significant difference. The maximal phase angle (Φ) for all other observers was set to 60°. However, all observers were also measured at Φ = 90° for 5 and 40 Hz presentation frequencies, again with no discernable difference.

The above experiments were carried out at 80% stimulus contrast. We repeated these experiments with five observers with an 18 cd/m² stimulus luminance, corresponding to 9% contrast, with presentation frequencies between 2.5 and 75 Hz. Thresholds for the two spatial configurations increased by about a factor of 2 [Fig. 3(a, b)]. The reduced contrast had no effect on perception of the asynchronously flashed stimuli [Fig. 3(c)].

**DISCUSSION**

While spatial displacements that are barely above the two-point acuity of about 1 min arc suffice to induce symmetry breaking in our bistable percept, introducing temporal offsets between the different parts of the triangle of up to 33 msec at frequencies between 10 and 75 Hz has only very little effect on whether or not an illusory triangle dominates. This is true at low as it is at high contrasts.
These results are compatible with earlier experiments showing that temporal synchrony has no effect on the rate of illusory conjunctions within a visual attentional paradigm (Keele, Cohen, Ivry, Liotti & Yee, 1988).

The experiments on temporal hyperacuity discussed above (Fahle, 1993) show that humans can reliably discriminate temporal offsets as small as 3–5 msec. Yet delays which are almost an order of magnitude larger do not interfere with the ability of the visual system to combine or bind disparate parts of a scene into a single object (see also Westheimer, 1990). Our results argue against the idea that the precise timing of events in the external world induces temporally synchronized neuronal responses in cortical neurons that are necessary for the figural binding of objects in pattern recognition, at least for the task to discriminate between figure and ground. Our results do not rule out the attractive notion that the internal code used by cortex for “binding” is contained in the detailed temporal correlation among cells (Milner, 1974; von der Malsburg, 1981; Gray et al., 1989; Crick & Koch, 1990; Kreiter & Singer, 1992). However, in figure–ground separation the internal code would bear no direct relationship to the timing of external events.

REFERENCES


Acknowledgements—This research is supported by the National Science Foundation, the Air Force Office of Scientific Research and the Deutsche Forschungsgemeinschaft (Fo 119/5-2, Heisenberg Program). C.K. would like to gratefully acknowledge the hospitality of the Institute for Theoretical Physics, Eidgenössische Technische Hochschule, Zürich, Switzerland during his sabatical. M.F. would like to thank M. Repnow and C. Winter for writing the computer code and H. Weller and A. Hildinger for technical assistance.