Attention and non-retinotopic feature integration

Laboratory of Psychophysics, Brain Mind Institute, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, & Laboratoire Psychologie de la Perception, Université Paris Descartes, Paris, France

Thomas U. Otto

Center for Neuro-Engineering and Cognitive Science, University of Houston, Houston, TX, USA, & Department of Electrical and Computer Engineering, University of Houston, Houston, TX, USA

Haluk Öğmen

Laboratory of Psychophysics, Brain Mind Institute, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

Michael H. Herzog

Features of moving objects are non-retinotopically integrated along their motion trajectories as demonstrated by a variety of recent studies. The mechanisms of non-retinotopic feature integration are largely unknown. Here, we investigated the role of attention in non-retinotopic feature integration by using the sequential metacontrast paradigm. A central line was offset either to the left or right. A sequence of flanking lines followed eliciting the percept of two diverging motion streams. Although the central line was invisible, its offset was perceived within the streams. Observers attended to one stream. If an offset was introduced to one of the flanking lines in the attended stream, this offset integrated with the central line offset. No integration occurred when the offset was in the non-attended stream. Here, we manipulated the allocation of attention by using an auditory cueing paradigm. First, we show that mandatory non-retinotopic integration occurred even when the cue came long after the motion sequence. Second, we used more than two streams of which two could merge. Offsets in different streams were integrated when the streams merged. However, offsets of one stream were not integrated when this stream had to be ignored. We propose a hierarchical two stage model, in which motion grouping determines mandatory feature integration while attention selects motion streams for optional feature integration.

Keywords: feature integration, non-retinotopic processing, metacontrast masking, motion grouping, attention, object


Introduction

Retinotopy refers to the spatial organization principle by which adjacent parts of the visual field are projected to neighboring neurons in the retina. This retinotopic organization is well preserved in many visual areas creating multiple two-dimensional maps of the visual field (Engel et al., 1994; Tootell et al., 1998). However, visual processing cannot always be retinotopic. For example, when we move our eyes, head, or body, the world appears as stable even though the retinal image undergoes dramatic changes. Non-retinotopic mechanisms are required to achieve this stability. For similar reasons, perception of moving objects also requires non-retinotopic processing. This has been demonstrated by a variety of recent studies showing that visual features are integrated along motion trajectories. These features include luminance (Shimozaki, Eckstein, & Thomas, 1999), form (Öğmen, Otto, & Herzog, 2006; Otto, Öğmen, & Herzog, 2006, 2009), color (Nishida, Watanabe, Kuriki, & Tokimoto, 2007; Watanabe & Nishida, 2007), size (Kawabe, 2008), the conjunction of features (Cavanagh, Holcombe, & Chou, 2008), and even motion itself (Boi, Öğmen, Krummenacher, Otto, & Herzog, 2009).

For example, in the sequential metacontrast paradigm, a single line is followed on either side by a sequence of flanking lines (Figure 1a; Otto et al., 2006). Observers perceive two diverging motion streams (Figure 1b). As in classical metacontrast masking (e.g., Alpern, 1953; Bachman, 1994; Breitmeyer & Öğmen, 2006; Picron, 1935; Stigler, 1910), the visibility of the central line is strongly suppressed by the flanking lines. Observers attend to one stream. When the central line is slightly offset, this central-offset is perceived within the attended motion stream (Figure 1b). This is an instance of non-retinotopic feature attribution because the offset of the central line is attributed to the flanking lines which are all non-offset. When one of the
whether motion grouping is the only factor for feature integration. Another factor may be attention, which has often been suggested to play a major role in establishing feature-object relations (e.g., Treisman, 1998; see also Footnote 1). The critical role of attention in our experiments is evident because the flank-offset was integrated with the central-offset only when it was presented in the attended stream but not when it was presented in the unattended stream.

Here, we investigated the effect of attention on non-retinotopic feature integration. By systematically varying when and where attention is allocated, we investigated whether non-retinotopic feature integration depends on the timing of (uni- vs. multi-focal) attention. A distinction that is critical in this context is whether or not attention can access the offsets prior to a feature integration stage. In the affirmative case, attention can read-out the offsets of the individual lines and, from the perspective of these processes, feature integration is optional (Figure 2d). In contrast, if attentional processes can only access the integrated offset attributed to an entire motion stream, from the perspective of these processes, feature integration is mandatory (Figure 2d). In two experiments, we manipulated the allocation of attention to different motion streams using a cueing paradigm in which an auditory tone defined the stream which had to be attended.

### General methods

#### Observers

Data were obtained from paid, naive observers and from one of the authors (T.O.). After the general purpose of the experiments was explained, observers signed informed consent. Observers were told that they could quit the experiment at any time they wished. For each observer, we determined the visual acuity by means of the Freiburg visual acuity test (Bach, 1996). To participate in the experiments, observers had to reach a value of 1.0 at least for one eye (corresponding to a Snellen fraction of 20/20).

#### Apparatus

Stimuli appeared on an X-Y-display (HP-1332A, Tektronix 606B, or 608) controlled by a PC via fast 16 bit D/A converters. Stimuli were composed of dots drawn with a dot pitch of 250–350 µm at a dot rate of 1 MHz. The dot pitch was selected so that dots slightly overlapped (i.e., the dot pitch was of the same magnitude as the dot size or line width). Refresh rate was 200 Hz. Stimulus luminance was 80 cd/m², as measured with a Minolta LS-100 luminance meter by means of a dot grid. The flanking lines is offset in addition, this flank-offset is integrated with the central-offset (Otto et al., 2006, 2009). However, this integration occurs only when the flank-offset is in the attended stream.

Three different but inter-related processes need to be distinguished: motion grouping, feature attribution, and feature integration. First, to perceive the single lines as one moving line, the visual system needs to establish a motion correspondence between the lines, i.e. motion grouping (Figure 2a). As the number of potential correspondences can be very large, establishing correspondences is often an ill-posed problem, the well known motion correspondence problem (e.g., Attneave, 1974; Dawson, 1991; Gepshtein & Kubovy, 2000; Kolers, 1972; Ullman, 1979). Second, feature attribution occurs when features of a single line are perceived at another line (Figure 2b). For example, the vernier offset of the central line is visible at the aligned lines of the motion streams (Figure 1). Finally, when a feature is attributed to a line that has a similar feature, will these features be perceived separately or will they be integrated (i.e. fused) into a single feature? We call the latter case, feature integration (Figure 2c). By integration, we mean a combination analogous to mathematical summation; for example, if the two offsets are of opposite direction, then they cancel each other.¹

In the last years, we have shown that non-retinotopic feature integration follows motion grouping (e.g., Ögmen et al., 2006; Otto et al., 2006, 2009). It remained unclear whether motion grouping is the only factor for feature integration. Another factor may be attention, which has often been suggested to play a major role in establishing feature-object relations (e.g., Treisman, 1998; see also Footnote 1). The critical role of attention in our experiments is evident because the flank-offset was integrated with the central-offset only when it was presented in the attended stream but not when it was presented in the unattended stream.

Here, we investigated the effect of attention on non-retinotopic feature integration. By systematically varying when and where attention is allocated, we investigated whether non-retinotopic feature integration depends on the timing of (uni- vs. multi-focal) attention. A distinction that is critical in this context is whether or not attention can access the offsets prior to a feature integration stage. In the affirmative case, attention can read-out the offsets of the individual lines and, from the perspective of these processes, feature integration is optional (Figure 2d). In contrast, if attentional processes can only access the integrated offset attributed to an entire motion stream, from the perspective of these processes, feature integration is mandatory (Figure 2d). In two experiments, we manipulated the allocation of attention to different motion streams using a cueing paradigm in which an auditory tone defined the stream which had to be attended.

### General methods

#### Observers

Data were obtained from paid, naive observers and from one of the authors (T.O.). After the general purpose of the experiments was explained, observers signed informed consent. Observers were told that they could quit the experiment at any time they wished. For each observer, we determined the visual acuity by means of the Freiburg visual acuity test (Bach, 1996). To participate in the experiments, observers had to reach a value of 1.0 at least for one eye (corresponding to a Snellen fraction of 20/20).

#### Apparatus

Stimuli appeared on an X-Y-display (HP-1332A, Tektronix 606B, or 608) controlled by a PC via fast 16 bit D/A converters. Stimuli were composed of dots drawn with a dot pitch of 250–350 µm at a dot rate of 1 MHz. The dot pitch was selected so that dots slightly overlapped (i.e., the dot pitch was of the same magnitude as the dot size or line width). Refresh rate was 200 Hz. Stimulus luminance was 80 cd/m², as measured with a Minolta LS-100 luminance meter by means of a dot grid. The
room was dimly illuminated (approximately 0.5 lux) and background luminance on the screen was below 1 cd/m². Viewing distance was 2 m.

**Stimuli**

We presented sequential metacontrast stimuli as introduced earlier (Otto et al., 2006, 2009). The sequence started with a central line consisting of two vertical segments of 10.0' (arcmin) length separated by a vertical gap of 1.0' (Figure 1a). Four pairs of flanking lines followed one after the other. The segment length for the first pair of flanking lines was 11.7' and increased progressively by 1.7' for the following lines. The centroid-to-centroid distance between the central line and the first flanking lines as well as between consecutive flanking lines was 3.3'.

![Figure 2](https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/933485/)  

**Figure 2.** (a) Motion grouping. Single lines appear as one line in apparent motion. Without motion grouping, a sequence of separate lines would be perceived. (b) Feature attribution. The offset of the first line is attributed to the moving line, which is perceived offset even though physically the later lines are aligned. (c) Feature integration. The right- and the left-offset of two lines are integrated in a manner analogous to mathematical integration (summation) and cancel out each other. (d) Attention may access the offsets of the single lines (dotted line) or only the integrated offset (solid line). In the former case, integration is optional, in the latter, it is mandatory.
Methods

Procedures

Each trial was initiated with four markers at the corners of the screen presented for 500 ms followed by a blank screen for another 500 ms. Then, the actual stimulus sequence was presented in the center of the screen. After stimulus presentation, a blank screen appeared and observers responded by a button press. No feedback was given. A new trial was initiated 1000 ms (500 ms in Experiment 2) after the observer gave a response.

We presented stimuli of different conditions in blocks of 120 trials randomly interleaved. For each observer, we presented each condition in a total of 160 trials distributed across four blocks. To reduce hysteresis, learning, or fatigue effects in the averaged data, we randomized the order of conditions/blocks across observers. Moreover, after every condition/block has been measured once, we reversed the order for the next set of conditions/blocks.

Experiment 1

Methods

To understand the role of attention in non-retinotopic feature integration, we varied the allocation of attention using a cueing paradigm. We asked observers to attend to one motion stream depending on an auditory cue (Figure 3a). In random order, a low frequency beep (500 Hz) cued the leftward and a high frequency beep (2000 Hz) the rightward motion stream. Block-by-block, the auditory cue was presented before, during, or after the sequential metacontrast sequence with one out of seven cue-stimulus onset asynchronies$^2$ (CSOA; test values: $-500, -50, 0, 80, 130, 180, and 500$ ms). In non-cued control conditions, observers attended always to the same motion stream as indicated at the beginning of a block.

Within the sequence, only the central line and/or penultimate flanking lines were offset. We employed three offset conditions (Figure 3b). In condition C, only the central line was offset. The direction of this central-offset was always the same as the direction of a notional reference offset which was randomly chosen on each trial. In condition C, only the penultimate flanking line of the cued stream was offset. This flank-offset was always in the opposite direction as the reference. In condition CF, both the central- and the flank-offset were presented. Because of the opposite offset directions, the flank-offset was to the left when the central-offset was to the right, and vice versa. We presented all combinations of cued motion stream (leftward, rightward) and offset condition (C, F, CF) randomly interleaved. Individual offset sizes ranged from 50" (arcsec) to 90" (mean: 68.3") for the central-offset and from 35° to 60° (mean: 49.2°) for the flank-offset (see General methods section).

We asked six observers to indicate the offset direction of the cued motion stream. In a binary forced-choice task, observers pressed the left/right button when they perceived the lower line segment displaced to the left/right with respect to the upper line segment. Observers were neither told that different offset conditions were presented nor that only a subset of lines in the display was offset (even if observers are aware of the paradigm, they cannot indicate which line in the attended stream is offset; Otto et al., 2006).

For each observer, we pooled the trials of each condition. For each trial, we recorded whether the observers’ response was in accordance with the direction of the notional reference offset. We computed percentages of responses in accordance with the reference. This percentage was above/below 50% when responses were dominated by an offset in the same/opposite direction as the reference. To ease intuition, we converted these percentages into dominance levels by subtracting 50%. With dominance levels, the sign indicates the direction of the dominating offset (central-offset: +; flank-offset: −). The absolute value indicates the strength of the corresponding dominance (ranging from 0% to 50%). For each condition, we computed the mean dominance level and the standard error of the mean (SEM) across observers. In addition, we recorded reaction times. We corrected for outliers by rejecting trials with (logarithmized) reaction times deviating by more than two-times the standard deviation from the mean. This procedure led to a rejection of less than 4% of the trials.

In a control experiment, we presented only two static lines simultaneously (see insert in Figure 5a). Each line consisted of two vertical segments of 10.0' length separated by a vertical gap of 1.0'. The centroid-to-centroid distance between the lines was 20.0'. The lines were presented for 20 ms. We asked observers to attend to one line depending on an auditory cue. We used CSOAs of $-500, -100,$
−50, 0, 50, 100, and 500 ms, respectively. The cued line was offset according to the individual threshold level. Individual offset sizes ranged from 8" to 24" (mean: 12.0°). The non-cued line had a 25% reduced offset size (i.e., ranging from 6" to 18", mean: 9.0°). The offset direction of the cued line was always in the same direction as the reference. The offset direction of the non-cued line was randomly in the same or opposite direction as the reference to control for attentional lapses.

**Results and discussion**

We analyzed performance as a function of CSOA. In condition C, performance was dominated by the central-offset (Figure 4a). With a CSOA of −500 ms, performance was similar to performance when observers attended always to the same stream in a block of trials (Figure 4a, broken green line). With increasing CSOA, performance deteriorated slightly. Similarly, in condition F, performance was dominated by the flank-offset and deteriorated slightly with increasing CSOA (as indicated by smaller absolute values of the dominance level). In both conditions, performance was similar whether the leftward or the rightward motion stream was cued (data not shown). In general, the decay of performance with increasing CSOAs seems to be similar to the decay of performance we found in the static control experiment (Figure 5a).

The central-offset was presented 130 ms before the flank-offset (see Figure 4, green and red bands, respectively). Despite this timing difference, the time course of performance decay was similar in conditions C and F. For
Figure 4. Cued attention. In all panels, the vertical gray band indicates the duration of the motion streams. Green and red bands indicate the duration of the central- and the flank-offset lines, respectively (Figure 3b). (a) Performance in conditions C and F as a function of cue-stimulus onset asynchrony (CSOA). The sign of the dominance level indicates whether the central- (+) or the flank-offset (−) dominates the responses. The absolute value reflects the strength of the corresponding dominance. For a CSOA of −500 ms, performance was similar to conditions when observers attended always the same stream, i.e. in non-cued conditions (dotted lines). With increasing CSOA, dominance decreased slightly but similarly for conditions C and F. (b) Performance in condition CF. For all CSOAs, performance was around 0%. The actual dominance level was well predicted by the sum of the dominance levels achieved in conditions C and F (C + F). Performance was similar when always the same stream was attended in a block of trials (dotted lines). (c) Lapses control. With the probe-offset in the non-cued stream, performance differs only slightly in same/opposite probe-offset trials (Figure 3c). This difference is close to the difference observed in a non-cued comparison condition (thin dotted lines) and much smaller compared to a comparison condition in which the stream with the probe-offset was attended (thin solid lines). (d) Reaction times. With a CSOA of −500 ms, reaction times were slightly longer than with one attended stream only (dotted lines). As the CSOA increased, reaction times increased. There was virtually no difference in reaction times for conditions C, F, and CF. Means and SEM for 6 observers.
statistical analysis, we compared the absolute values of the dominance level using a repeated measures ANOVA with CSOA and offset condition as factors. We found a main effect of CSOA ($F(6, 30) = 3.86, p = 0.006$) but neither a main effect of offset condition ($p = 0.772$) nor an interaction effect ($p = 0.745$). Because the decay seems not to be specific to the timing (condition C vs. F) or stimulus type (sequential metacontrast vs. static control), we suggest that it results from a stimulus-independent central mechanism (see also, Plomp, Mercier, Otto, Blanke, & Herzog, 2009).

Next, we analyzed performance when both the central- and the flank-offset were presented. In condition CF, we found that the dominance level was always around 0% indicating that the central- and flank-offset cancelled each other (Figure 4b, CF). Interestingly, for all CSOAs, performance for condition CF was well predicted by the sum of the dominance levels achieved in conditions C and F (Figure 4b, C + F; see also, Otto et al., 2009). This finding indicates that the integration of the central- and the flank-offset is not changed by cueing.

Next, we determined the contribution of the non-cued stream to performance. In randomly interleaved half of the trials, we presented a third offset in the non-cued stream (Figure 3c). This probe-offset was either in the same or in the opposite direction as the reference. Hence, if the non-cued stream contributes to performance, we expect a systematic difference in performance between “same” and “opposite” probe-offset trials (as in a comparison condition in which observers attended always to the stream with the probe-offset; see Figure 4c, thin solid lines). If the non-cued stream does not contribute to performance, a difference should not occur (as in a comparison condition in which observers attended always to the stream without the probe-offset; see Figure 4c, thin broken lines). With the auditory cue, performance in “same” and “opposite” trials was rather similar to the latter comparison condition. Smaller differences occurred only for CSOAs of −50, 0, and 500 ms (Figure 4c). This indicates that the non-cued stream is largely ignored and that only a few attentional lapses (i.e., observers may have erroneously selected the non-cued stream) occurred when the cue was presented either shortly before or long after the motion streams.

The analysis of reaction times showed that observers held their decision until the cue was presented (Figure 4d; for static lines, see Figure 5b). Interestingly, even when observers attended always to the same stream, there is virtually no difference in reaction times in the conditions C, F, and CF (Figure 4d, dotted lines). This indicates that observers did not take advantage of the fact that the central-offset was presented 130 ms before the flank-offset.

To summarize, we varied the allocation of attention by an auditory cue. When the cue was presented before the visual stimulus, observers could focus attention on the cued stream from its onset to its offset. This corresponds to the case where uni-focal attention is allocated to the stream of interest. When the cue was delayed, observers had to divide attention across streams until the cue indicated the stream to attend. This corresponds to

Figure 5. Control experiment. An auditory cue indicated whether a left (L) or a right static line (R) had to be attended. The cued line was offset in the same direction as the reference. The non-cued line was offset randomly in the same (as in R) or in the opposite direction as the reference (as in L). We varied the onset of the cue with respect to the onset of the lines (on-screen as indicated by the gray band).

(a) Performance. With a CSOA of −500 ms, the dominance level was similar to conditions when observers attended always to the same line in a block of trials (dotted line). Dominance decreased slightly with increasing CSOA. For all CSOAs, we found no performance differences whether the left or right line was cued (data not shown). The same held true when we compared trials with the non-cued line in the same or in the opposite direction as the reference (data not shown). (b) Reaction times increased the later the cue was presented. Mean and SEM for 6 observers.
the case where attention is divided among streams for part of, or for the entire duration of the motion stream depending on the timing of the cue. We found a general decay of performance with increasing CSOAs, which was similar to the decay of performance in a control experiment with static lines. Interestingly, the decay showed a similar time course for conditions C and F although the SOA between central- and flank-offset was 130 ms. Taken together, these observations indicate that the decay does not depend on the specifics of stimulus conditions and, thus, results from a stimulus-independent central mechanism (see also, Plomp et al., 2009). Neither the timing nor the distribution of attention had a specific effect on feature integration. In fact, despite the general decay, performance in condition CF was always well predicted by the performance levels in conditions C and F. Hence, non-retinotopic feature integration within a motion stream was similar when attention was allocated exclusively to one stream or had to be divided across streams. These findings indicate that attention cannot directly access single lines and mandatory feature integration occurs within the attended motion stream (see also, Otto et al., 2006, 2009). In contrast, even if attention was divided across streams, the probe-offset presented in the un-cued stream was largely ignored indicating that there is no integration across streams. This issue will be further investigated in the next experiment.

**Experiment 2**

The results of Experiment 1 show that uni-focal attention is not necessary for non-retinotopic feature integration. Moreover, if attention is divided across streams, offsets presented in two different streams are not integrated even if the cue comes after the motion streams. This implies that the offsets in each motion stream need to be stored in memory. By using the principle of Occam’s razor, we considered the simplest memory mechanism that can implement mandatory integration. In computing technology, in order to avoid time consuming and expensive read/write operations to the main memory, an accumulator memory is used to store intermediate results of operations. This is like in a relay in athletics, each single line within the motion stream receives, as a baton, an offset value from the preceding line. This offset is then integrated with the line’s own offset. The integrated offset is then passed on to the next line, and so on. At some point, a motion stream might be cued and attention reads out the integrated offset stored in the accumulator memory at the line presented last. If two motion streams coincide in one location, such a model would predict that the offsets of both streams are integrated in a mandatory fashion and that attention cannot modulate the outcome. This hypothesis is tested in the following.

**Methods**

We presented two sequential metacontrast sequences simultaneously (Figures 6a and 6b). The sequences were presented next to each other so that two of the flanking lines coincided in the last frame. Hence, the last frame contained three lines only. The distance between the two lines in the first frame was 26.7’.

Within the stream, only the lines in the first frame were offset. The direction of these offsets (left, right) was selected with respect to a notional reference offset as in the first experiment. We employed three offset conditions (Figure 6c). In condition A, only the left line in the first frame was offset. This offset A was always in the same direction as the reference. In condition B, only the right line was offset. This offset B was always in the opposite direction as the reference. In condition AB, both lines were offset. Because of the opposite offset directions with respect to the reference, offset B was to the left when offset A was to the right, and vice versa. We presented the three offset conditions (A, B, and AB) randomly interleaved within a block of trials. We used the same offset size for offset A and B. In the first part of the experiment, individual offset sizes ranged from 60” to 100” (mean: 75.0") and, in the second part, from 60” to 120” (mean: 86.0”).

**Results and discussion**

As with one sequence only (Figure 1), the visibility of the two lines in the first frame is strongly reduced. Block-by-block, we asked observers to indicate the offset direction of one stream. In the first part, we asked observers to indicate the offset of the last central line. Both offsets yielded similar performance (Figure 7a, A and B). When both offsets were presented, the dominance of either offset was strongly reduced indicating that the two offsets were integrated (Figure 7a, AB; there is only a slight dominance of offset A). This finding is in agreement with the accumulator/relay model of memory.

Remarkably, because of the random presentation of offset conditions, observers did not know whether offset A or B was presented and could not focus attention on the corresponding line. Hence, uni-focal attention on one line in the first frame is not necessary for non-retinotopic feature attribution.

When the leftmost line was attended in condition A, performance was dominated by offset A (Figure 7b, Left). In condition B, performance was close to 0% indicating that offset B is not attributed to the leftmost motion stream. In condition AB, performance was similar to condition A. Hence, the two offsets were not integrated. Analogous results held true when the rightmost line was attended (Figure 7b, Right).

As a control experiment, we removed the last frame containing three lines so that the two central motion streams
did not merge. We asked observers to attend either to the left or the right of the two center lines (Figure 7c; Center-L and Center-R). Performance was similar to the conditions when observers attended the leftmost and rightmost line, respectively (Figure 7b). Hence, the center line in the last frame seems to be necessary for the integration of the offsets A and B.

According to the accumulator/relay model, integration is mandatory when two streams merge at one retinotopic location. To test this prediction, we presented the same sequences as in the first part (the last lines were presented for 40 ms to strengthen the percept of two motion streams merging). We asked five new observers to attend to the center line (as in the first part) or to the center line and to attend selectively to only one of the two streams.

When the center line was attended, performance was similar to the first part (Figure 8a, see also see Figure 7a). Performance was strongly changed when only one of the merging streams was attended (Figure 8b). When the stream coming from the left was attended, offset A dominated performance and offset B contributed only marginally (Figure 8b, Center-L). Analogously, when the stream coming from the right was attended, responses were dominated by offset B and offset A contributed only marginally (Figure 8b, Center-R). Hence, depending on the allocation of attention, observers reported selectively offset A or B. This shows that integration across streams is not mandatory even if streams merge in one retinotopic location.

With respect to the accumulator/relay model, one explanation might be that observers could access either stream before the two streams merged. To control for this, we presented an auditory cue with a CSOA of 500 ms (i.e., more than 300 ms after the two streams merged). Performance did not change qualitatively (Figure 8c). The result is similar compared to when observers attended only one stream in a block of trials (Figure 8b). We interpret this finding as evidence against mandatory integration like in a accumulator/relay. Unlike the accumulator/relay memory, which retains only information about the most recent outcome of a mandatory integration, characteristics of the two motion streams must have been stored in separate memories to allow attention to modulate the outcome of feature integration. In other words, feature integration across streams is optional.

**Discussion**

When observers attend to one motion stream, non-retinotopic feature integration occurs for the lines within this attended motion stream (Otto et al., 2006). We proposed that motion grouping determines the integration
of offsets. To shed light on the role of attention in non-retinotopic feature integration, we manipulated the allocation of attention across motion streams in two experiments.

In Experiment 1, we varied the allocation of attention to one motion stream by an auditory cue. Our results show that feature integration occurs even when the cue came after the stimulus sequence (Figure 4). With post-cueing, uni-focal attention cannot be exclusively allocated to one stream. This finding indicates that feature integration is pre-attentive or attention works either in parallel by multiple “spotlights” or by a rapidly alternating spotlight (VanRullen, Carlson, & Cavanagh, 2007). Future research will test integration with more motion streams that have to be tracked (Alvarez & Cavanagh, 2005; Alvarez & Franconeri, 2007; Howard & Holcombe, 2008; Pylyshyn & Storm, 1988; Shooner, Tripathy, Bedell, & Ögmen, 2010; Tripathy, Narasimhan, & Barrett, 2007; Yantis, 1992).

When cue time increased, performance decayed similarly in the central- and flank-offset conditions C and F (Figure 4a). If the allocation of attention was locked to the onset of the offset line (see also, Figure 5), a shift in the decay function would be expected because the flank-offset was presented 130 ms later than the central-offset. It seems that allocation of attention relates to the entire motion stream rather than to individual lines and when they are presented in the stream.

This hypothesis is further supported by the fact that observers cannot cognitively access the single lines within a motion stream. We have previously shown that observers cannot discriminate whether the central line or the flanking line was offset even though observers can discriminate well the direction of the offset perceived in the motion stream (Otto et al., 2006). This finding provides strong evidence for mandatory offset integration within a motion stream (at least within the spatio-temporal range covered by our stimuli). On the contrary, if we present the lines with the central- and the flank-offset but without the flanking lines in-between, the motion percept is strongly reduced and observers perceive two rather distinct flashes of lines. Consequently, offsets are not integrated (Otto et al., 2009; see also Cavanagh et al., 2008). Finally, we have shown that non-retinotopic feature integration and attribution are changed by stimulus manipulations that are well known to change the motion grouping (Ögmen et al., 2006). Hence, we find strong evidence that non-retinotopic feature integration within a single stream is guided by motion grouping and not by attention.

In Experiment 2, we investigated non-retinotopic feature integration across motion streams. We considered a simple and efficient model that can implement mandatory integration. The offsets of single lines might be integrated...
like in a relay. At each time point, the offset stored at one retinotopic location is given by the offset of the present line and the offset remapped from the location of the preceding line. An advantage of such a relay model is that, for each motion stream, only one offset value has to be stored. Such a memory, called accumulator, is widely used in computing technology due to its simplicity and efficiency. However, such accumulator/relay models are challenged by our results. As shown in Figure 7, the offsets of two streams are combined when merging into one line. Still, feature integration is strongly influenced by the allocation of attention to one stream even if a cue is presented 300 ms after stimulus disappearance (Figure 8c). This shows that integration of offsets presented within different motion streams (even if these converge in one retinotopic location) is not mandatory but can be modulated depending on attention and task demands.

For this reason, we propose that non-retinotopic feature integration requires the maintenance of information about motion streams in memory and the ability of attention to select and transfer this information to decision processes (see also, Kahneman, Treisman, & Gibbs, 1992; Pylyshyn, 1989; Yantis, 1992). At one extreme, one can assume that all lines are stored in visual short-term memory for temporally extended processing and their offsets can be selected by attention according to the task instruction. However, as discussed before, experimental evidence suggests that attention cannot individually access lines that are grouped within one stream. For example, attention cannot “ignore” a flank-offset presented in the attended stream. For this reason, we propose the following hierarchical two stage “grouping-attention” model of non-retinotopic feature integration (see also, Scharnowski, Hermens, Kammer, Öğmen, & Herzog, 2007). Motion grouping mechanisms “sort” single elements into various basic groups (i.e., the motion streams in our experiments). When grouping of the single elements is ambiguous, certain elements may belong to more than one group (e.g., the central line may belong to the left and the right motion stream, see Figure 1). Attention cannot “access” the single lines in the basic groups. However, attention can operate on these groups and combine them into higher-order groups depending on task demands. Features are then integrated according to the selection of groups. For example, the offsets of the two central streams are integrated (by default) when attention is not focused selectively on either stream (Figure 7). When observers are instructed to attend to one of the central streams only, the offset in the other stream does not, or only marginally, contribute to performance (Figure 8). Hence, attention guides optional feature integration across streams.

Figure 8. (a) We repeated the experiment with attention to the central line. Results are similar to Figure 7a. (b) We asked observers to focus on the stream coming from the left (Center-L) or right (Center-R). When the motion stream coming from the left was attended, offset A dominated and offset B contributed only marginally. When the stream coming from the right was attended, we found the inverse. (c) Post-cueing. An auditory cue with a CSOA of 500 ms indicated whether the stream coming from the left or right had to be attended. Performance is very similar to (b) except for a small decay in performance like in Experiment 1. Mean and SEM of 5 observers.
Our results are in accordance with earlier studies on static displays. For example, it has been shown that the resolution of attention is much coarser than spatial resolution itself (He, Cavanagh, & Intriligator, 1996). Also in these static cases, grouping seems to prevent attention to “access” features of individual elements when closely nearby, possibly because of spatial grouping of single elements into arrays of elements (see also, Parkes, Lund, Angelucci, Solomón, & Morgan, 2001).

Summary

We propose that non-retinotopic feature integration is guided by motion grouping which establishes a continuity of objects across space and time. Within a motion stream, integration of features is mandatory and does not require uni-focal attention. Integrated features are stored in a short-term memory labeled not by retinotopic locations but by the motion streams themselves, which form the elementary units for attentional selection. Attention can select amongst motion streams for further processing.

Acknowledgments

Thomas U. Otto was supported by the Swiss National Science Foundation project “The Dynamics of Feature Integration.” H. Ögmen was supported in part by award number R01 EY018165 from NIH. The content is solely the responsibility of the authors and does not necessarily represent the official views of NIH. We thank Marc Repnow for technical support and Jeroen van Boxtel for helpful comments on an earlier version of the manuscript.

Commercial relationships: none.
Corresponding author: Thomas U. Otto.
Email: thomas.otto@parisdescartes.fr.
Address: Laboratoire Psychologie de la Perception, Université Paris Descartes, 45 rue des Saints-Pères, 75006 Paris, France.

Footnotes

1Note that our use of the term feature integration is different from the terminology used by Treisman and colleagues. In their use, feature integration refers to the process whereby different features, such as color and shape, are labeled to belong to the same object.

2CSOAs represent physical asynchronies between the onset of the auditory cue and the onset of the visual stimulus, and not neural asynchronies. This is because the transmission delays of sensory stimuli depend on stimulus characteristics and the latencies associated with auditory and visual stimuli are different. Hence, a CSOA of 0 ms, for example, does not mean that the auditory and the visual stimuli are synchronous in the cortex.

3While the latency of an auditory stimulus is in general shorter than the latency of a visual stimulus, a CSOA of 500 ms corresponds to a time interval between the offset of the visual stimulus and the onset of the auditory cue of 310 ms or 290 ms depending on the experiment, a value long enough for the auditory cue signal to reach the cortex after the offset of the visual stimulus.

References


