Perception of successive targets presented in invariant-item streams

Talis Bachmann *, Pilleriin Sikka

Estonian Center of Behavioral and Health Sciences, Perception and Consciousness Research Group, University of Tartu, Kaarli puiestee 3, Tallinn 10119, Estonia

Received 8 December 2004; received in revised form 23 February 2005; accepted 28 February 2005
Available online 3 May 2005

Abstract

When two successive, spatially overlapping, targets (S1 and S2) are presented on a blank background, S2 typically dominates in explicit perception. We tested whether S2 dominance is also found for the conditions of presenting S1 and S2 in a stream of irrelevant objects. Successive target letters were presented within a stream of invariant stimulus items (capital Is). The stream items were presented either as a perceptually continuous object where both type and token appeared invariant (60-Hz stream) or as a flickering stream of successive replicas of the invariant stationary object where the type appeared invariant but the token appearance seemed intermittent (20-Hz condition). Compared to the control condition where targets were presented on a blank background we found that (1) recognition rate was lower for targets embedded in a perceptually continuous type-and-token object (60 Hz), but was unchanged for targets in a perceptually flickering sequence of the invariant-object tokens (20 Hz); (2) S1 recognition rate was higher compared to S2 recognition rate within the first epoch of stream (0–150 ms) while within the later stream-epochs S2 dominated over S1 as usual; (3) the overall difference in recognition rates between S1 and S2 was decreased. The results are discussed in the theoretical context of visual masking and attentional blink.

© 2005 Elsevier B.V. All rights reserved.

PsycINFO classification: 2323; 2346

* Corresponding author. Tel.: +372 6 271891.
E-mail address: talis.b@lawinst.ee (T. Bachmann).

0001-6918/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.actpsy.2005.02.007
1. Introduction

If two different target–stimulus objects are presented from the same spatial location in rapid succession and the time intervals are about 40–100 ms, the second one typically dominates in explicit perception (Bachmann, 1989; Bachmann & Allik, 1976; Michaels & Turvey, 1979). On the other hand, if two different targets (S1 and S2) are presented within a stream of varying objects, all from the same location, the following target may be perceived much worse compared to the preceding target. This occurs with inter-target stimulus onset asynchronies (SOAs) of about 200–400 ms and provided that S1 is successfully recognised, a phenomenon called attentional blink (Chun & Potter, 1995; Raymond, Shapiro, & Arnell, 1992). If S1 and S2 are presented on a blank background, attentional blink does not occur. Because in the RSVP streams used in the studies of the attentional blink the distractor items are mutually different successive objects it is not clear whether identity-processing of the variable stream objects or streaming per se cause the difference in relative S1/S2 recognition rate compared to S1–S2 presentation in isolation. As a first step in approaching this problem, we decided to test if streaming of stimulus items that creates a temporal stimulation context for target perception that is similar to RSVP could be the condition that changes typical masking functions where S2 dominates S1. These are obtained when the stimuli for perception are presented on a blank background (e.g., Bachmann & Allik, 1976; Michaels & Turvey, 1979). The changed mutual-masking functions of S2/S1 relative perception in the within-stream temporal context may be similar to the attentional blink functions found with RSVP streams (e.g., Raymond et al., 1992).

The obvious way to disentangle streaming per se from the variable identity-processing of different successive objects consists in presenting formally invariant replicas of the same stream-object in rapid succession. In this case it becomes possible to use pure streaming where object tokens are presented multiple times as temporally individuated items but object type remains invariant. This makes processing for identity or token individuation in the identity domain redundant. (See Kahneman & Treisman, 1984; Kanwisher, 1987, 2001, for both the more detailed analysis of the type-and-token distinction and the importance of understanding the mechanisms of specification of the incoming sensory signals as discrete objects.) Awareness of the form attribute of the irrelevant stream items should remain invariant but individuation of the corresponding perceptual information as pertaining to a distinct event should vary.

In order to obtain a variable temporal individuation of the invariant-type tokens, the frequency of stream presentation should allow distinct perceptual intermittency of the successive objects, for instance at 20 Hz. In order to obtain perceptual invariance both at the token and the type identity levels, the frequency of stream presen-
tation should be increased so as to lead to the perception of a steady, unchanging, invariant object (e.g., 60-Hz presentation).

The aim of the present study is to compare elementary perceptual recognition of successive targets in three different conditions: (1) when target stimuli are presented within a stream of temporally individuated but formally invariant items (a “proto-stream” of object tokens without a change of successive objects’ shape so that variability of identity-related individuation is eliminated) or in isolation, (2) when target stimuli are temporally presented within a perceptually continuous, unchanging and steady object, or (3) when spatially overlapping targets are presented on a blank background as in typical mutual masking. As a result, we will obtain data showing how streamness per se, allowing stream-token individuation in time and avoiding variability in the individuation of stream item types can change perceptual microgenesis of objects.

In order to explore the time course of object microgenesis in stream we use the experimental paradigm of mutual masking (successive form recognition) combined with a technique that is similar to RSVP but where the successive stimulus items that form the stream are invariant. In successive form recognition two mutually different successive targets (S1 and S2) are presented for identification with varying stimulus onset asynchrony (SOA) and they act as forward and backward masks to each other (Bachmann & Allik, 1976; Michaels & Turvey, 1979). This technique has enabled to observe the time course of perceptual recognition in a quite precise way (see Fig. 1). It takes about 150 ms for the S1 to escape from backward masking. Also, with intermediate SOAs around 40–70 ms S2 strongly dominates perception so as if “substituting” or “replacing” S1 in explicit representation (Bachmann & Allik, 1976; Enns & DiLollo, 2000). With longer SOAs around 150–200 ms perception of S1 and S2 becomes equally good at a high level. We are interested whether this set of effects may

![Fig. 1. Typical results of the mutual masking (successive form recognition) studies where it takes about 150 ms SOA for the first object (S1) to become free from backward masking and where the second object (S2) dominates in perception, “substituting” or “replacing” S1 in explicit perception at intermediate SOAs. (Adapted from Bachmann and Allik, 1976; Michaels and Turvey, 1979.)](image-url)
be modified by the within-stream presentation in the conditions where targets are clearly different from the stream items that remain invariant.

In the experiment reported here we have presented varying S1 and S2 both in isolation on a blank field (as in typical mutual-masking studies) and within the two types of streamed uninformative perceptual items. We varied the frequency with which the stream items were presented between 20 Hz and 60 Hz. Thus a flickering token condition and a non-flickering (perceptually invariant) token condition can be guaranteed. The refresh cycle of the light background of the dark stimuli was always set at 60 Hz, but the dark stimuli representing stream items that appear on the light background were presented with a varying frequency. With 20 Hz, all subjects distinctly perceived flicker of the stream objects and the stream therefore had the explicit “stream quality”. With 60 Hz, the physical stream of the invariant stream-objects appeared perceptually as a steady, continuously presented object. Because the temporal frequency and luminance of the luminous flickering background was always invariant, the integrative luminance-masking effects from the background possibly exerted on target identification should have been similar between the conditions of different frequencies of the stream item presentation.

We put forward the following hypotheses: 1. When targets are presented within streams, S2, correct identification rate should be higher than S1 correct identification rate, consistent with typical mutual-masking regularities (the effect of order of targets should be significant and in the direction favouring S2 as the better perceived stimulus). 2. Because the invariant items in the 60-Hz and 20-Hz streams are individuated as types in the same invariant way and because the power of luminance-masking of targets from the flickering luminous background is similar between the two frequency conditions, similar results conforming to the first hypothesis should be obtained in both the frequency conditions. 3. Because stream items exert an extra load on visual processing, we expect stronger masking in stream than in the control condition where S1 and S2 are presented on a blank background. 4. We expect decrease of stimulus identification with increase in SOA because of decrease in the effect of masking. 5. We hypothesize that S2 will be perceived better than S1 in all time epochs of the stream because all stream items are identical, more stream items as backward masks follow S1 than S2, all time intervals between successive stimuli are identical, and the luminous background of the stimulation has the same refresh rate throughout all the experimental conditions.

2. Experiment

2.1. Method

2.1.1. Participants

Six subjects aged between 18 and 51 years participated in the experiment. All participants had normal or corrected-to-normal vision. One participant was one of the authors, the others were recruited for the first time and were naïve as to the theoretical context of the study.
2.1.2. Apparatus and stimuli

The stimuli were presented on a PC screen and the refresh rate of the display was 60 Hz (synchronisation with the experimental program-dependent stimulus timing was executed in the DOS regime). All stimulus items were depicted as dark letters (4 cd/m²) on a light background (100 cd/m²). The streams consisted of spatially invariant, successive, identical letters (I), forming a stream with a total duration equal to 1200 ms (including two targets). One frame of a stream item lasted for 17 ms. In order to obtain the 60-Hz stream the stream items were presented in each successive frame throughout the whole duration of the stream. Perceptually, all subjects experienced the 60-Hz physical stream of stimuli as a non-flickering, steady object (the letter I). In order to obtain the 20-Hz presentation rate of the stream, two empty frames with blank screen were inserted into the stream after each stream item (or target).

The vertical dimension of the stream item I amounted to about 0.4° of the visual angle, observed from the subject’s point of view. Inserted at unpredictable temporal positions within the stream were two temporally separated target stimuli (S1 and S2) presented at the same location as the stream items in the main experimental conditions. In the two display-frames, the stream items (I) were substituted with the targets. All targets were presented for 17 ms each. They also subtended 0.4° vertically. On each trial, two mutually different targets were chosen randomly from the pool of alternative target letters A, B, C, D, E, F, G, H, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, and Y. The font was chosen so that if superimposed spatially with the stream item I, the target letters were easily discriminable. In the 60-Hz condition, inter-target SOA was varied randomly between 17 ms, 33 ms, 50 ms, 67 ms, 83 ms, 100 ms, 117 ms, 133 ms, and 150 ms. Because in the 20-Hz condition the number of possible inter-target SOA values was smaller than the number of possible inter-target SOA values in the 60-Hz condition, SOAs in the 20-Hz condition were varied among 50 ms, 100 ms, and 150 ms. The number of trials with each particular SOA in the 20-Hz condition was thus proportionally increased in comparison with the number of trials used with that of SOA in the 60-Hz condition. In the control condition, targets were presented without visible stream items (stream items were defined as blank white in the respective computer program in order to obtain zero luminance- and colour-contrast between the background and the stream-stimulus areas). Except this, all other conditions were the same for the main experimental conditions (60-Hz and 20-Hz streams) and the control condition.

2.1.3. Design and procedure

A repeated measures multifactor design was used. There were five independent variables of interest: streaming of stimuli (two levels—presentation of targets in stream and control condition without stream), frequency of stream item presentation (two levels—60-Hz stream and 20-Hz stream), temporal order of target stimuli (two levels—S1 and S2), SOA (nine or three levels), and epoch of stream, i.e., time interval from the beginning of stream until S1 presentation (four levels—50–150 ms, 250–350 ms, 550–650 ms, 950–1050 ms). The two stream-frequency conditions were applied in blocks. In the first type of block where 60-Hz frequency was used the
physically streamed input appeared as a perceptually steady, invariant item. In the second type of block where 20-Hz frequency was used the streamed input appeared as a temporally intermittent, but formally invariant stimulus-item. The order of running the types of frequency-blocks was counterbalanced between and within the participants. The number of control trials without stream was equal to the number of streamed presentation trials. Order of targets contributed to each trial and SOA was varied randomly; thus, there were \((4 \times 2 =) 8\) blocks of 400 trials applied with counterbalanced order within each frequency-block. Subjects was a random factor.

The task was to identify both targets (S1 and S2) and type in the responses by a computer keyboard. Subjects were requested to respond always with two responses, guessing if unsure. They sat about 60 cm away from the computer monitor. Trials were initiated by the Enter key, followed by the appearance of the fixation dot (four screen pixels arranged as a square). In the streamed presentation condition, a visible stream of Is including the targets followed the fixation dot (see Fig. 2); in the control condition, no visible stream of Is appeared after the fixation dot and targets were presented on a blank background. The order in which targets were reported was unimportant; thus, subjects were allowed to type responses in any order they liked. Responses (two different letters) were recorded after a press of the Enter key. The next press of the Enter key initiated the following trial. No feedback about the correctness of the response was provided in the experiment. Before recording experimental data, 100 practice trials were administered to each subject with feedback about correct responses provided.

2.2. Results and discussion

According to the first hypothesis we expected that S2 correct identification rate should be higher than S1 correct identification rate. An analysis of variance revealed

![Fig. 2. A schematic diagram of the stimulus presentation used in our experiments. Uninformative, invariant items (I) form a stream of spatially overlapping successive objects. Within this stream, two successive targets (S1 and S2) are presented with varying SOAs. (Actually, the number of stream items is considerably larger than depicted in this illustration.)](image-url)
that the overall rate of correct identification was significantly higher for S2 than for S1 \((F(1,240) = 71.325, p < 0.0001)\); correct target identification percentages were 68% and 59% (in the 60-Hz condition), and 84% and 76% (in the 20-Hz condition), respectively. Our first hypothesis is supported. This regularity repeats the well-known facts from mutual masking on a blank background where S2 typically has an advantage over S1 (Bachmann & Allik, 1976; Michaels & Turvey, 1979). For example, in Fig. 3(a) and (b), recognition rates of S1 and S2 as a function of SOA are drawn for the control condition and the in-stream condition with 60-Hz streams (averaged over all epochs of stream). In these conditions sufficient number of SOAs were used in order to allow drawing a usual masking function. A qualitatively similar picture appeared in the 20-Hz stream condition with only 3 SOAs.

The second hypothesis predicted equality of masking in the two frequency conditions because of invariance of the stream items and flickering luminous background between these conditions. A comparison of the effects of the two frequency conditions showed that target perception was significantly better in the 20-Hz condition than in the 60-Hz condition \((F(1,240) = 64.737, p < 0.0001)\). A separate post hoc
analysis confirmed this ($F(1,5) = 22.4, p < 0.01$). Thus, the second hypothesis was not satisfied. For S1, correct identification rate in the 20-Hz stream condition was 73%, but in the 60-Hz stream it was only 42%. (The respective percentage values for S2 were 78% and 46%.) This result suggests relative facilitation of perception (or release from masking) by a stream with perceptual flicker compared to the condition with a steady, continuous object impression created by the stream items. The putative relation of this result to the differences in the ways the object tokens are individuated should be explored in future experiments.

Because the difference between 60-Hz and 20-Hz conditions was virtually equal for S1 and S2 then it could be in principle also attributed to integrative masking of targets by stream items. (Due to the shorter SOAs, masking may have been stronger with 60-Hz stream than with 20-Hz stream.) This should leave interruptive or substitutional masking effects between targets and stream items out of consideration. If interruption or substitution were the case, S2 perception should have suffered relatively more from stream presentation than S1 perception because the stream items that follow S2 should in turn substitute S2 in visible representation. However, the integrative masking explanation is also suspect because the refresh rate of the luminous background that reduces the contrast of the dark target stimuli was set invariantly at 60 Hz in both, the 20-Hz and the 60-Hz stream conditions.

The rate of correct target identification was significantly higher in the control condition than in the main conditions where targets were presented within streams ($F(1,240) = 208.432, p < 0.0001$). The third hypothesis is therefore supported. Within-stream presentation of targets seems to add power to masking. But there was a significant interaction between streaming condition (control vs main) and frequency of stream items ($F(1,240) = 95.413, p < 0.0001$). Masking of targets was stronger in the 60-Hz condition and much weaker in the 20-Hz condition. The validity of the second hypothesis is also suspect—we see that masking level depends on the frequency of stream. An analysis of variance carried out exclusively for the 60-Hz condition showed that the rate of correct target identification was significantly higher in the control condition than in the main condition where targets were presented within the steady-appearing streams ($F(1,5) = 10.6, p < 0.01$). S1 was correctly identified in the in-stream and control conditions on 42% and on 76% of trials, respectively. Respective percentages for S2 were 46% and 90%. Because luminance-masking of targets from the light background of the display (which had invariant 60-Hz frequency throughout the whole study) was equal in the streamed and control conditions, this difference has to be related either to metacontrast-like masking of targets by the stream items or to the higher-level processing influences related to competition between targets as objects and stream items as objects because both types of items have to be individuated. The masking effect exerted on target identification by the invariant, but otherwise uninformative stream items in the 60-Hz presentation condition outweighed the possible facilitative effects of the stream as such on the processing of the in-stream items in comparison with the isolated items (the facilitative effect was introduced by Bachmann, Luiga, Põder, & Kalev, 2003; Bachmann & Põder, 2001).
The facilitation effect in Bachmann et al. studies appeared as a decrease in visual latency of the in-stream targets compared to the replicas of the targets presented in isolation. Therefore, even if there may have been some effects of stream on the speed of processing, these were either outweighed by the between-object masking effects or localised at some definite epochs of stream (and not distributed uniformly over the whole stream).

An analysis of variance carried out exclusively for the 20-Hz condition showed that the rate of correct target identification was not significantly different in the control conditions and in the conditions where targets were presented within streams ($F(1, 5) = 1.3$, $p > 0.05$). This result is different from what was found in the 60-Hz condition. It appears that the 20-Hz frequency of stream presentation, creating token intermittency, enables to perceive targets as effectively as in the control condition where targets were presented on a blank background. Obviously, streamness as such and perceptual flicker with a 20-Hz stream, which are characteristic of the typical RSVP streams, do not change the expression of mutual masking compared to standard masking conditions without a stream. But because in the 60-Hz streams masking was definitely stronger than in the control conditions we should find an explanation for this discrepancy.

First, there may be a stronger between-object masking in the 60-Hz condition where the steady-appearing stream item strongly competes with the target-items for encoding. Alternatively, it may be that in the 60-Hz condition S1 more strongly competes with S2 compared to this competition in the 20-Hz condition. Second, the steady non-flickering appearance of the stream item I in the 60-Hz condition may have made it more difficult to individuate target tokens that were temporally embedded within the continuous-appearing stream-object. Further studies should help answer whether the same form of between-object masking or the difficulties in token individuation (or both) can explain the difference between the 60-Hz and the 20-Hz conditions.
20-Hz conditions. Intuitively, perceptual individuation of the successive items in time as a microgenetic Gestalt-building process seems to be the key process here. In the flickering stream, successive individuation processes begin ab ovo with each newly appearing successive object (or object replica); therefore there is less between-object competition. Otherwise, were the preceding stream-object representation not terminated, this competition would have been effective. The newly appearing target object is free from rivalry with the preceding stream-object by virtue of avoiding representational conflict. This is because the stream-object has been perceptually terminated before the appearance of the target and thus it cannot contribute to competition any more. In a steady-object stream, a newly appearing target must overcome a strong and stable representation of the individuated stream item that still occupies the same spatial location. By the rule of consistency in the operation of the internal object-processing algorithms, two different objects cannot occupy the same spatial location at once. Therefore, a strong mutual competition at the level of objects is inevitable.

Consistently with the fourth hypothesis, stimulus identification rate increased with increase in SOA ($F(2,240) = 9.137, p < 0.0001$). (Only SOA values common to both frequency conditions included in the analysis.) Release from mutual-masking effects takes place also when target objects are presented within the streams of invariant distractor items.

There was no significant main effect of the epoch of stream within which targets were presented ($F(3,240) = 0.611, p = 0.608$). This result seems to support our fifth hypothesis. However, we found a significant interaction between target order (whether S1 or S2) and epoch ($F(3,240) = 3.927, p < 0.009$), with the direction of the effect referring to some unusual properties of mutual masking if targets are presented early within a stream. The unusual picture of S1/S2 relative rate of correct perception was found both in the 60-Hz stream condition and in the 20-Hz stream condition. In the first epoch S1 dominated over S2, but with later epochs this tendency reversed showing the usual dominance of S2 over S1 (see Figs. 4 and 5).
60-Hz stream, in the first epoch S1 was perceived correctly in 42% of trials and S2 in 38% of trials, but with later epochs this tendency reversed (in epoch IV, 46% correct for S1 and 54% correct for S2). With 20-Hz stream, while in the epochs II, III, and IV S2 dominated over S1, in the first epoch S1 was identified better than S2 (respective percentages—78% and 71%). (Fig. 5 shows S1 and S2 recognition rates as a function of epoch of stream in the 20-Hz condition.) These results include aspects that are exactly opposite to what was hypothesized in our first hypothesis and also contradict the fifth hypothesis.

There seem to be several possibilities of explanation that need not be mutually exclusive: (1) a relative in-stream facilitation of S1 processing in the first epoch, found specifically when stream items preceded S1 for no more than 50–100 ms (e.g., Fig. 4); (2) a more stronger masking of S2 than that of S1 especially in the first epoch of the stream (in these conditions, S1 correct identification exceeded S2 correct identification by 5–14%, while in the epochs II–IV, S2 slightly dominated S1); (3) masking was stronger at the onset of the stream-object because it has to be individuated at the beginning of its processing and the first epoch provides stronger inter-object competition (between target and stream objects) compared to the later epochs where type individuation for the stream item is obviously redundant. The puzzling aspect is that for some reason this putative inter-object competition affects S2 processing more than S1 processing. Future studies should specifically explore the possible reasons of this kind of difference between S1 and S2 as the competitors for the stream item in terms of their type individuation.

The first one of the above-mentioned possibilities is consistent with the fact that the largest S1 facilitation, measured in terms of speed of perception, has also been found specifically in the earliest epoch of stream compared to the later epochs where

![Fig. 6. Percent correct S1 recognition in the 60-Hz and 20-Hz conditions drawn separately for different epochs of S1 presentation, compared to no-stream control condition.](image-url)
facilitation was still present, however obviously diminished (Bachmann & Oja, 2003). The in-stream facilitation was obtained quickly at the onset of stream, but did not accumulate slowly along the extended time intervals within the stream. Thus it may be possible that relative demasking of S1 is due to its relatively stronger facilitation by the first items of stream. In order to be sure that there is some form of facilitation that outweighs masking we should refer to the conditions where in-stream masking is not stronger than masking in isolated presentation conditions.

At the same time, S1 should be also similarly facilitated in comparison with S2. The results obtained in the 20-Hz condition support this possibility: a separate analysis of variance showed that the rate of correct target identification across SOAs was not significantly different in the control condition and in the condition where targets were presented within 20-Hz streams ($F(1,5) = 1.3, p > 0.05$).

In the 60-Hz condition, possibly due to the increase in the effect of masking in that condition (compared to the control condition), the relative advantage of S1 over S2 in the first epoch of stream could have been emphasised because masking of S2 became stronger. However, with regard to the 20-Hz condition where masking was not stronger than in the control condition, we should more seriously consider the possibility of S1 facilitation in stream, provided that the targets are presented in the first epoch of stream. The relative facilitation of S1 was strong especially if stream items preceded S1 for no more than 50–100 ms.

Although S1 is relatively facilitated when presented in the first epoch of stream, its backward masking by S2 is nevertheless evident. This is so especially in the 60-Hz condition. Curiously, however, in the 20-Hz condition S1 is perceived at the level equal to the level found in the control condition without the stream (see Fig. 6). The 20-Hz temporal frequency is quite fast in terms of what has been typically used in the RSVP experiments. But as the stream consists of formally invariant items, it does not add to the masking power.

3. General discussion

In the standard mutual masking or successive form recognition experiments where two targets, S1 and S2, are presented on a blank background, S2 strongly dominates in correct perception (Bachmann, 1989; Bachmann & Allik, 1976; Michaels & Turvey, 1979). In the attentional blink, to the contrary, S1 dominates and S2 perception is impaired (Raymond et al., 1992). In the experiment reported in this study we tested whether targets in a stream of invariant-type items are perceived similarly to the targets presented in isolation, i.e., S2 dominating over S1. We also tested whether perceptual flicker of the stream, allowing temporal token individuation, could bring in differences for targets’ perception compared to the targets’ perception in a stream that was perceived as a steady, continuous, non-flickering item. Our main findings were: 1. S2 correct perception rate in general exceeded that of S1. 2. S2 correct perception rate was closer to S1 correct perception rate if S2 and S1 were presented within 60-Hz physical stream producing a perceptually steady object (compared to S1–S2 presentation without a stream, where S2 strongly dominated). 3. Within the
first epoch of stream and with short SOAs, the relative rates of correct perception of S1 and S2 were reversed compared to typical results of mutual masking: in these specific conditions, S1 was perceived better than S2. 4. The rate of correct target recognition in the 20-Hz stream condition was at the same level as in the control condition without a stream, but recognition rate decreased considerably in the 60-Hz condition. Perceptual flicker and streamness as such are not the factors that would lead to increase in mutual masking within streams; however the steady continuous perceptual object formed from the successive invariant stream item(s) provides a difficult context for targets to be individuated and correctly perceived. It is easier to individuate targets if invariant stream items flicker with successive SOAs that are similar to inter-target SOAs.

In general, with type-wise invariant stream items and short inter-target SOAs in the range of 50–100 ms, the dynamics of successive form perception is qualitatively more similar to the perception in typical mutual masking than to the RSVP-effects such as attentional blink (e.g., Chun & Potter, 1995; Raymond et al., 1992). There is one exception though—in the first epoch of stream and with relatively shorter SOAs between targets (e.g., 50 ms), S1 perception prevails over S2 perception. Differently from the in-stream facilitation found in the flash-lag experiments, the overall correct recognition rate of the present in-stream targets was lower than the overall correct recognition rate of the isolated targets. (In the flash-lag effect, an isolated target object that is flashed simultaneously with another object that appears within a stream of stimulation is perceived as lagging behind the in-stream object. As revealed by the temporal order judgement task, targets in stream were perceived as appearing faster than isolated targets—Bachmann et al., 2003; Bachmann & Pöder, 2001; see also Scharlau, 2002.) On the other hand, within the first epoch of a stream (at about 50–150 ms after stream onset), S1 correct perception level exceeded S2 correct perception level (see Figs. 4 and 5). In terms of in-stream facilitation as measured by correct perceptual recognition of masked visual objects, primarily the initial epoch of the stream after its onset (up to about 150 ms) provides conditions for the relative perceptual facilitation of S1. In a stream, and within its initial epoch, S1 is capable of overcoming the typically mandatory strong backward masking exerted on it by S2.

We can outline two main possibilities in order to explain the in-stream relative facilitation for S1 within the first epoch of stream. First, there may be a proactive facilitation of S1 by the onset of the first items of stream so that the slow non-specific facilitatory modulation evoked by stream onset becomes effective at the time when the S1-specific signals are encoded by cortical driver neurons. As a result, the signal-to-noise ratio of S1-related activity in the cortex is enhanced and perceptual saliency of S1 is increased to the extent that backward masking by S2 decreases. (For the details of this explanation, consult perceptual retouch theory in Bachmann, 1999.) If S1 and S2 are presented in isolation, S1 itself as the first-appearing stimulus initiates a slow facilitatory modulation that becomes effectively applied onto the cortical areas and, coincidentally, this happens no sooner than the S2 specific signals are encoded by the cortical driver neurons. This is what leads to the typical S2 dominance over S1. But this conjecture forces us to modify the
perceptual retouch theory so that the mechanisms that boost non-specific facilitation after the appearance of the new stimuli have to be (1) especially sensitive to the onsets of new stimulation and (2) liable to a relatively fast exhaustion which therefore explains why the S1 facilitation effect subsides in the later stream epochs.

Our data showing S1 relative recovery as compared to S2 perception, specifically within the first epoch of stream presentation (50–150 ms after stream onset), suggests that the putative frequency synchronisation that should facilitate integration of target features into perceptual objects may have been strongest within the first epoch of stimulation. Actually, as found by Herrmann and Mecklinger (2001), the strongest evoked gamma activity in response to visual forms was revealed exactly within this time interval, 50–150 ms after stimulus onset. In the conditions of streamed stimulation, stimulation onset seems to be the most significant factor for possible stimulus facilitation in comparison with other factors such as frequency of stimulation or type of stimulus.

Secondly, there are several attention-related processing accounts that can also be accommodated to explain facilitation of S1 after stream onset. In the asynchronous updating model (Scharlau & Neumann, 2003), it is assumed that stimuli are processed by two processes, feature coding and attentional allocation. Feature coding is preattentive and undergoes continuous fast updating. Attentional allocation is needed for the transfer of stimulus information to a consciously available internal model.

Attentional allocation is slower than preattentive coding. If a stream is begun, both coding and attentional allocation are triggered, but attentional allocation takes more time and will be optimised at the moment when S1 is presented. Therefore, attention will be best applied exactly when the stimulation stream is updated by the S1 signals, which will be the main contents of the internal model, partly escaping from backward masking caused by S2. This is because attentional allocation had been prepared by the stream onset. Quite similarly, the temporal profile model of processing presumes optimisation of attention with a certain temporal delay (Stelmach & Herdman, 1991). It may be of no mere coincidence that the notion of updating and the notion of token individuation (Kanwisher, 1987) both help to understand object discrimination within the streams of rapidly alternating, spatially overlapping objects. If updating is very easy due to the token-type invariance, as was the case with our invariant-item streams, the token individuation is also easy. This allows S1–S2 relative perception to follow the usual dynamics as found in standard mutual masking.

There was an important difference between our experiments and the standard attentional blink studies (e.g., Chun & Potter, 1995; Raymond et al., 1992): we used stream items that were invariant replicas of the same stimulus and at the same time always visually different from the targets. In the study by Wagemans, Verhulst, and De Winter (2003) attentional blink was studied in the conditions where the shapes of the non-target items were also distinctly different from the target shapes, somewhat like in our study. They found that even in the conditions where the visual categories of the forms of the first target and the second target are qualitatively different (letters and silhouettes, respectively), there was virtually no attentional blink. Although in
our experiment shorter times between targets were used and the categories of target forms for S1 and S2 were the same, we also found that contrary to the typical attentional blink, S2 perception was not compromised. Instead, S2 usually dominated over S1, even if presented within a stream. This was a prevailing result, except for when the targets were presented within the first epoch of stream. It seems that if target set does not change and if stream items can be easily discriminated from targets, typical inter-object masking with S2 dominance over S1 prevails.

In a different way, Kellie and Shapiro (2004) have shown that attentional blink considerably diminishes if some object-file continuity is present between the stream items. The stream items in our study were actually as continuous as possible because of their formal and spatial invariance. It is thus by no means surprising that attentional blink was not found. What was surprising though was the fact that the streams with more conspicuous continuity actually created more difficulties for successful target perception than the streams with less continuity. But maybe continuity of distractor tokens combined with invariance of the distractor type provides a more difficult background for target individuation than the continually changing distractor type?

Perception of objects within streamed sensory input is in several ways different from the perception of objects that are presented in isolation. These effects, however, cannot be limited to the inhibitory or interference effects that are typically occurring in the attentional blink with varying distractor items. Token individuation difficulties apparent even with invariant stream items and in-stream relative facilitation effects in the beginning of a stream should be considered and taken into account as well.

Acknowledgments

Support from Estonian Science Foundation grants #4967 and #5778 is gratefully acknowledged. We highly appreciate the help from Endel Pöder and Iiris Luiga at various stages of this research. We are also grateful to Ingrid Scharlau and an anonymous reviewer for very useful suggestions and constructive criticism. We are very much pleased with the patience and a constructive attitude of Johan Wagemans who provided his help far beyond that is usually typical for journal editors.

References


