Temporal windows and visual capacity

Andreas Wutz

PhD thesis

CIMeC - Doctoral school in Cognitive
and Brain Sciences

26th cycle

2013

Supervision:

David Melcher
Alfonso Caramazza

UNIVERSITÀ DEGLI STUDI
DI TRENTO
CIMeC - Center for Mind/Brain Sciences

CIMeC
Center for Mind/Brain Sciences

Acknowledgments

Thanks to my dad who gave me strength. Thanks to my mum who gave me confidence. Thanks to David who started to believe in me and with whom it was an honor to think with. Thanks to Doris who brought me here. Thanks to my friends who directed me elsewhere. Thanks to Jen who showed me the way. Thanks to Prof. Caramazza, Nathan and Christoph who taught me so much. Thanks to Dea who helped me to finish this thing.

Signs of gratitude also to Vladimir Horowitz and Frédéric Chopin, the Talking Heads, Daft Punk, the XX and Bon Iver for all these hours.

The research was supported in part by a European Research Council Grant (Grant agreement n. 313658); by the Italian Ministry of Universities and Research (MIUR), Project PRIN 2009; and by the Fondazione Cassa di Risparmio di Trento e Rovereto.

Table of contents

List of Figures

List of Tables

1. Temporal windows and visual capacity

The present work aims to arrive at a functional understanding of the visual computations within one perceptual instant of time. When perceiving the external world, we feel surrounded by a sensory environment that extends continuously into both space and time. However, the content of conscious thought consists of coherent scenes containing a limited number of discrete objects as an invariant percept within one particular instance of time. Here, vision is investigated as a process that extracts spatio-temporally invariant information about the physical world and at the same time integrates the current perceptual representation into a dynamic stream of visual impressions.

1.1 Temporal windows in vision

1.1.1 Evidence for 'perceptual moments' in vision

Phenomenologically, the visual environment appears temporally continuous and also classical physics are built on the premise of a continuous flow of time:

"Absolute, true and mathematical time by itself and from its own nature flows equably without relation to anything external" (Newton, 1689, p. 6).

However, whether visual perception operates upon a continuous signal or on a discrete sequence of events is still subject to debate (VanRullen& Koch, 2003). A first definition of such a perceptual moment was introduced by von Baer as 'the longest possible time interval for an organism still to be considered as a 'time point'' (von Baer, 1864). Shortly after, Ernst Mach estimated the duration of this interval. Temporal duration shorter than 30-40 ms cannot be discriminated (Mach, 1965). Intervals below that threshold appear to have no duration at all and are experienced as time points.

These early observations have led the way to the concept that sensory

information arriving within these perceptual moments is collapsed into a single percept and these temporal frames are simultaneously segmented into discrete informational chunks. Such temporal windows in visual processing have been a core feature within several variants of a 'perceptual moment hypothesis', in which the visual system quantizes incoming sensory information into temporally successive packages of finite duration (Stroud, 1956; White, 1963; Shallice, 1964; Allport, 1968; for review see Pöppel, 2009). Further evidence for discrete processing comes from periodicities in response histograms of choice reaction times or eye movement frequencies (Figure 1.1).

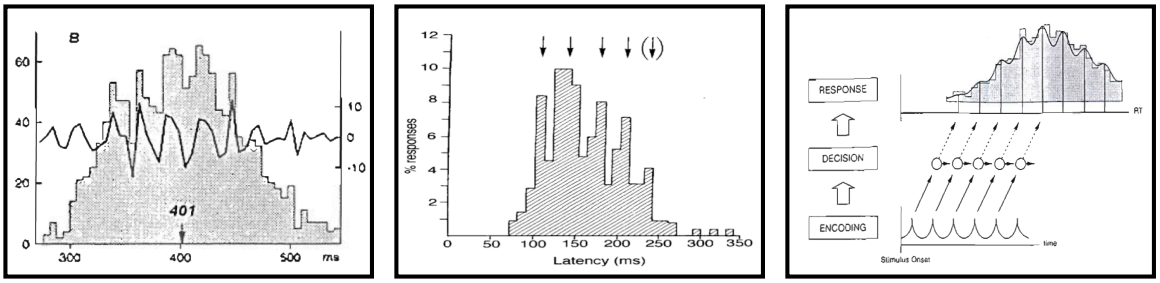


Figure 1.1 Oscillations in behavior

Periodicities in manual reaction times (left and right panel) and eye movement frequencies (middle panel) are indicative of discrete processing cycles (Figures adapted from Jokeit, 1990; Pöppel & Logothetis, 1986 and Dehaene, 1993)

Whereas such 'behavioral oscillations' are indicative of discrete informational chunks on the decision or motor level, they do not necessarily imply discrete processing on the level of sensory encoding (Figure 1.1 right panel). More direct support for such 'perceptual framing' is revealed in electrophysiological studies that show that the temporal relation between sensory stimulation and neural processes alters the perceptual outcome. Along these lines, psychophysical threshold estimates have been shown to vary with the phase of ongoing oscillatory activity (Mathewson, Gratton, Fabiani, Beck & Ro, 2009; Busch, Dubois & VanRullen, 2009). Moreover, perceived simultaneity and sequentiality of apparent motion percepts depend on the phase of the

occipital alpha rhythm (Varela, Toro, Roy & Schwartz, 1981; Gho & Varela, 1988; Figure 1.2).

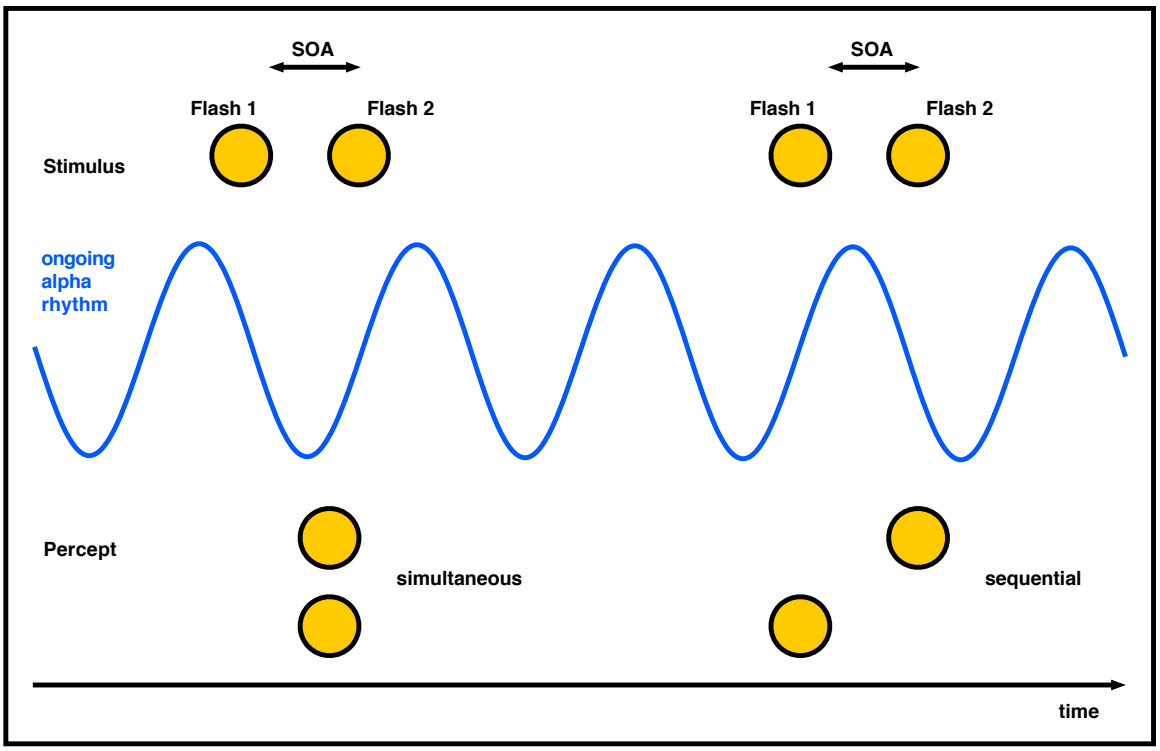


Figure 1.2 Perceptual framing and neuronal oscillations
Two successive events can be perceived as simultaneous or sequential depending on their temporal relation to the phase of ongoing oscillatory activity (Figure adapted from Varela, 1981; Gho & Varela, 1988 and VanRullen & Koch, 2003)

The majority of approaches towards perceptual moments reviewed here can be categorized to the concept of 'cinematographic vision' (Sacks, 1992): vision as consisting of a rapidly flickering series of stills, as in a slide show.

Indeed, such oscillatory biasing between active and inactive states has been in particular linked to ongoing background processes of temporal attention (VanRullen, Carlson & Cavanagh, 2007; Hanslmayr, Gross, Klimesch, & Shapiro, 2011). In this way attentional sampling might alter transient sensory detection (Landau & Fries, 2012). Visual information as a result of such camera-like passive registration is considered to

be virtually unlimited in capacity, but also fragile and easily over-written by subsequent input (Wundt, 1989; Sperling, 1960; Breitmeyer & Ögmen, 2006). Such short-lived and fragile snapshots can be used to compute global summary statistics of the raw sensory image. For example, the average size of a set of objects can be computed even when the display properties change continuously (Albrecht & Scholl, 2010). Since the time course of oscillations is inherently predictable, periodic sampling of such global scene properties might provide long-term stability and continuity of 'the gist' of the sensory image. Indeed, such 'virtual representations' have been conceptualized as a layout upon which "focused attention" extracts individual objects in a scene (Rensink, 2000) and they could help to integrate sensory information globally across different frames of reference (Corbett & Melcher, under review).

1.1.2 Functional characteristics of 'perceptual moments'

The present work, however, augments the camera-metaphor of vision by focusing on the computations within the snapshot. This includes exploring a second main characteristic of brain oscillations apart from their temporal periodicity: robust phase synchronization to transient input (Buszaki & Draguhn, 2004). Resets might either be induced top-down via saccadic eye-movements or shifts in attentional focus or evoked by real-world transitions (i.e. stimulus onset). In fact, evoked responses to successfully detected and entirely missed stimuli differ extensively (Busch et al., 2009). Temporal attention has classically been described as involving an initial transient component (Nakayama & Mackeben, 1989) and also oscillatory variations in susceptibility to sensory input have to be understood in relation to an initial reset event (Landau & Fries, 2012).

Visual operations immediately after transient sampling are particularly crucial since the parallel computed, but fragile sensory image is translated into a capacity-

limited and stable set of individual entities (Sperling, 1960; 1963). Within the first unit cycle from phase reset, persisting information from the sampled sensory image is segmented into invariant object representations (Kahneman, Triesman & Gibbs, 1992; Rensink, 2000). Invariant perceptual form has to be segregated virtually in real-time in order to encompass the problems of a retinotopic snapshot representation within a dynamically changing sensory surrounding (Ögmen & Herzog, 2010).

In contrast to this demand for high temporal resolution, coherent integration of visual information over time relies on synchrony of convergent inputs. "Synchrony is defined by the temporal window within which some trace of an earlier event is retained, which then alters the response to a subsequent event" (Buszaki & Draguhn, 2004; p. 2). In order to perceive visual continuity new information arriving within this critical window of sensory accumulation should change the current perceptual representation. The present work examines this balance between sensitivity to change and reading-out invariant perceptual form within a rapid temporal window in response to transient reset as a potential 'perceptual moment' in vision.

1.1.3 Stability of information and sensitivity to change within a 'perceptual moment'

The perception of a stable and coherent external world comprehensively relies on parsing of the sensory environment into individual informational units (Spelke, 1988). Humans do not perceive the physical world as a collection of unrelated features, but as organized and structured percepts of objects and scenes (Tipper, Brehaut & Driver, 1990; Kahneman, Treisman & Gibbs, 1992; Baylis & Driver, 1993; Scholl, Pylyshyn & Feldman, 2001). In this way, objects (or more abstract demonstrative pointers to them) form fundamental perceptual units of information, providing a link between sensation and cognition (Pylyshyn, 2001; 2007).

Moreover, stability and continuity of the current perceptual representation is preserved despite of dynamic changes of the sensory input. During normal vision sensory information undergoes drastic and complex changes due to either real-world transitions or internal shifts in processing focus. Saccadic eye-movements create large discontinuities in the flow of visual information at a rate of 3-4 alterations per second (for review see Rayner, 1998). Likewise, the sensory image can change rapidly in dynamic environments. A briefly presented stimulus remains visible for another 100 - 120 ms after its onset, a phenomenon called visual persistence (Haber & Standing, 1970; Coltheart, 1980; Di Lollo, 1980; explained in more detail below). If the visual system operates upon retinotopic snapshot-like images, fast changing or moving objects would be expected to appear smeared along the motion path (Figure 1.3).

Figure 1.3 Illustration of the 'problem of moving ghosts'
Whereas static objects (like the telephone cell to the right) appear clear and sharp, dynamic objects are highly blurred (like the bus to the left).

Stationary objects remain long enough on a well-defined location in the image, so that its associated features can be firmly attached to this location. The visual persistence of objects in motion, however, would be spread along successive locations of its motion path ('the problem of moving ghosts', Ögmen, 2007; Ögmen & Herzog, 2010). In normal vision, however, motion blur is only seldom perceived (Ramachandran, Rao & Vidyasagar, 1974; Burr, 1980; Hogben & Di Lollo, 1985).

Perceiving visual stability and continuity therefore imposes a fundamental functional dichotomy on visual processing: the construction of stable and coherent objects and scenes while also remaining sensitive to new information in the sensory image with high temporal resolution (Melcher, 2011). Given that sensory input arrives continuously, the visual system must mediate between stable and flexible representations virtually in real time.

1.2 Units of information in vision

1.2.1 Object files and object individuation

Extracting objects from sensory input is called object individuation and involves selecting features from a crowded scene, binding them into a unitary representation and individuating this spatiotemporal unit from other individuals in the image (Treisman & Gelade, 1980; Xu & Chun, 2009).

Object representations at this stage are suggested to be coarse and contain only minimal feature information. In fact such individual entities do not necessarily provide information about object identity, but can be regarded as a spatio-temporal placeholder of the object in focus. Several theoretical, psychophysical and neuroimaging studies have emphasized the computational importance and necessity of such incremental object representations in intermediate-level vision named visual indexes (FINSTs

(acronym for FINgers of INSTantiation), Pylyshyn, 1989), proto objects (Rensink, 2000) or object files (Kahneman, Treisman & Gibbs, 1992; Xu & Chun, 2006). Reduced to its minimum such an object file can be defined as an entity whose recent spatio-temporal history can be reviewed and therefore still can be referred to as the same entity despite of changes in its location over time (Kahneman et al., 1992).

The concept of object files can also be very intuitively understood as the percept evoking the following well-known statement:

"It is a bird! It is a plane! … It is Superman!"

Whereas identity information about the object in focus is not known, the observer perceives a unitary representation that is individual, discrete and well differentiated from its background (Smith, 1998). Object representations on this level of processing are commonly measured with an enumeration task that solely requires knowing whether an object is an individual rather than its identity. The number of open object files serves as a measure of stability and coherence of the current perceptual representation, since only successful individuation allows for further and more elaborate processing.

In a dynamic surrounding, non-retinotopic perceptual form must be computed to encompass the 'moving ghost problem' (Ögmen, 2007). In this context, individuation as a measure of local saliency has been hypothesized as a metric within a dynamic reference frame (Ögmen & Herzog, 2010). Since such a dynamic reference is based on motion segmentation, the number of individual relative motion vectors defines the informational resolution of the representation in time and space.

1.2.2 Visual capacity limits

It has long been noted that individuation is limited in capacity: we can quickly and effortlessly perceive that there are exactly three items but not that there are exactly thirty items (Jevons, 1871; Figure 1.4).

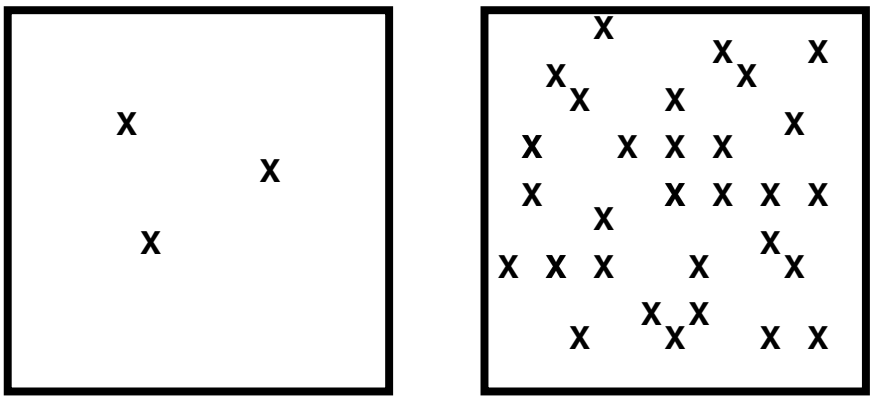


Figure 1.4 Stimuli in an enumeration task that measures object individuation.
Three items in the left panel can be simultaneously apprehended. Individuation of 30 items in the right panel requires successive perceptual steps (counting)

Enumeration is equally quick, accurate and effortless within a narrow range of one to four objects. Performance for set sizes exceeding this range, however, deteriorates more with every additional item to be enumerated (Figure 1.5).

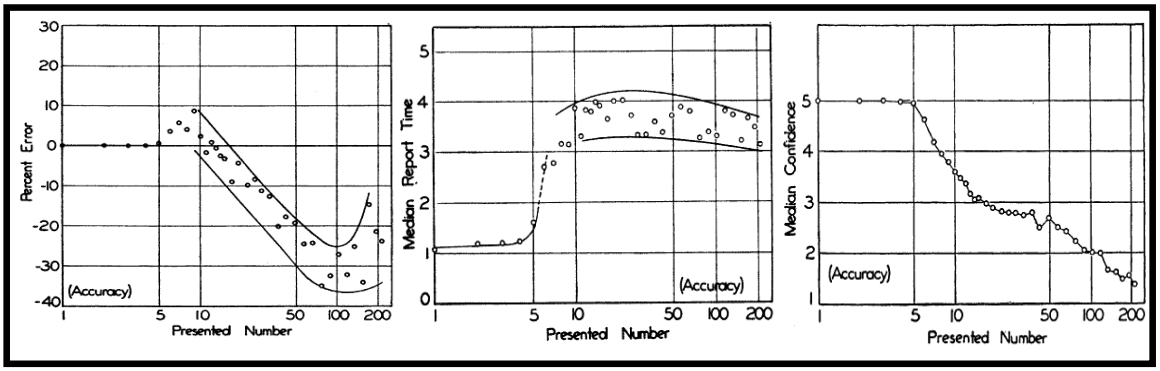


Figure 1.5 Enumeration performance as a function of number of presented items
The left panel shows the proportions of errors, the middle panel shows reaction time and the right panel shows reported confidence in judgment on a 1-5 scale (Figures adapted

from Kaufmann, Lord, Reese & Volkmann, 1949).

Whereas enumeration of five or more items has to rely on serial and time-consuming counting or imprecise estimation, smaller numerosities are presumably simultaneously apprehended by a qualitatively distinct mechanism known as “subitizing” (Kaufmann, Lord, Reese & Volkmann, 1949). Conceptually, subitizing determines the informational capacity of vision within one processing iteration.

Some of these individuated object files are elaborated subsequently during object identification. It is at this stage that identity information becomes available to the observer and the content of the object files can be consolidated into durable and reportable representations in visual working memory (Xu & Chun, 2006). As individuation precedes identification, the capacity of the latter has its upper bound in the limit of the former (Melcher & Piazza, 2011; Piazza, Fumarola, Chinello & Melcher, 2011, Dempere-Marco, Melcher & Deco, 2012; compare Figure 1.5 left panel and Figure 1.6).

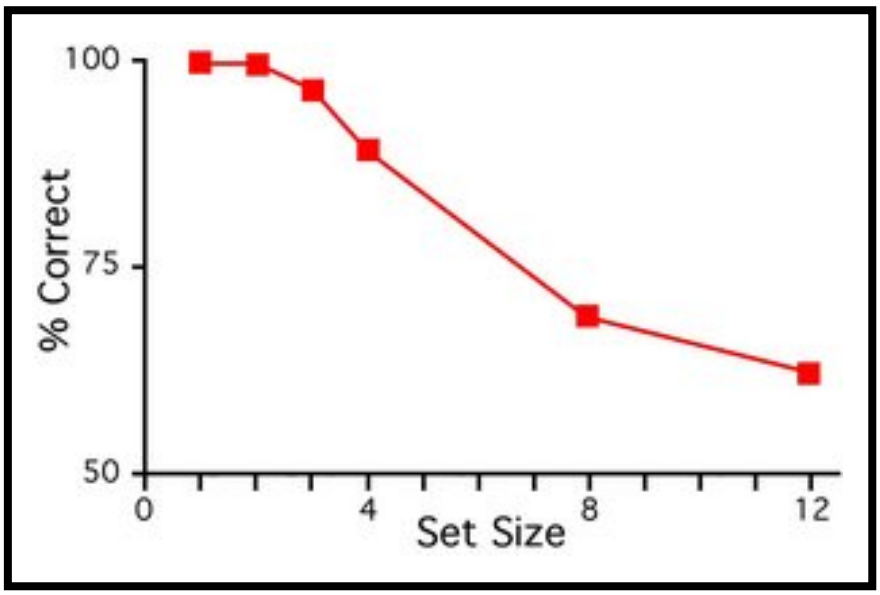


Figure 1.6 Change detection performance as a function of number of presented items Visual working memory capacity does not exceed the subitizing limit (Figure adapted from Luck & Vogel, 1997)

Although human cognition is remarkably powerful, its online workspace, working memory, appears to be highly limited in the number of informational units it processes (Sperling, 1960; Luck & Vogel, 1997; Cowan, 2000). It is interesting to note that this capacity is linked to cognitive abilities in general. For example, inter-individual variability in measures of fluid intelligence and capacity estimates are highly correlated (Engle, Tuholski, Laughlin & Conway, 1999; Cowan et al., 2005; Fukuda, Vogel, Mayr & Awh, 2010) and reduced capacity is often found in patients with neuropsychiatric disorders (Karatekin & Asarnow, 1998; Lee, Cowan, Vogel, Rolan, Valle-Inclán & Hackley, 2010). In light of its importance for cognitive functioning, the search for the root of this capacity limitation is fundamental to the study of cognition.

Object individuation appears to be the initial (but along with identification and consolidation not the only) bottleneck in visual processing from an unlimited in capacity, but fragile, purely bottom-up and in parallel computed sensory representation (iconic memory: Sperling, 1960, 1963; Neisser, 1967) to a capacity limited, durable and cognitively structured visual store (visual short-term memory: Sperling, 1960, 1963; Phillips & Baddeley, 1971). In addition to its influence on visual working memory and visual cognition in general, it is at the stage of object individuation that stability of the current perceptual representation is achieved through the establishment of spatio-temporally invariant object files. There are a number of competing theories for why subitizing, and individuation in general, is limited to sets of only about three or four items (for review, see Piazza, Fumarola, Chinello & Melcher, 2011), but it is widely assumed that this limit arises from a uniform process in one instant. In other words, the informational capacity of visual processing is thought to extend simultaneously to a small set of up to four informational units or object files through the continuous passage of time.

1.2.3 Temporal dynamics of individuation: Visual persistence and temporal integration

Previous examinations on the temporal dynamics of individuation, however, might have under-sampled its time scale, since information about the sensory image might be available longer than its actual physical presence (Sperling, 1960). Therefore a description of subitizing as purely instantaneous, mapping the apprehension of information on perceptual time with a transient impulse δ-function (Dirac, 1958), might be inadequate.

Following stimulus onset a briefly presented visual display has a limited perceptual persistence during which time it may be processed and categorized (Wundt, 1899; Sperling, 1960; Loftus, Duncan & Gehrig, 1992). This interval has been described as a sensory integration period over time. When two visual stimuli are presented in rapid succession (within 100 – 150 ms), their trailing visual persistence are partly integrated into a single percept. Ongoing neural processes still active from the first stimulus can dramatically reduce the visible persistence of the second stimulus, a phenomenon called masking by integration of contours (Di Lollo, 1980). Integration masking occurs when target and mask information are combined together, as a consequence of the imprecise temporal resolution of the visual system (Scheerer, 1973 a; Enns & Di Lollo, 2000; Breitmeyer & Ögmen, 2006). This forward masking manipulation makes it possible to quantitatively change the duration of visual persistence (and iconic memory access) by varying the onset asynchrony between the first and second display (Figure 1.7).

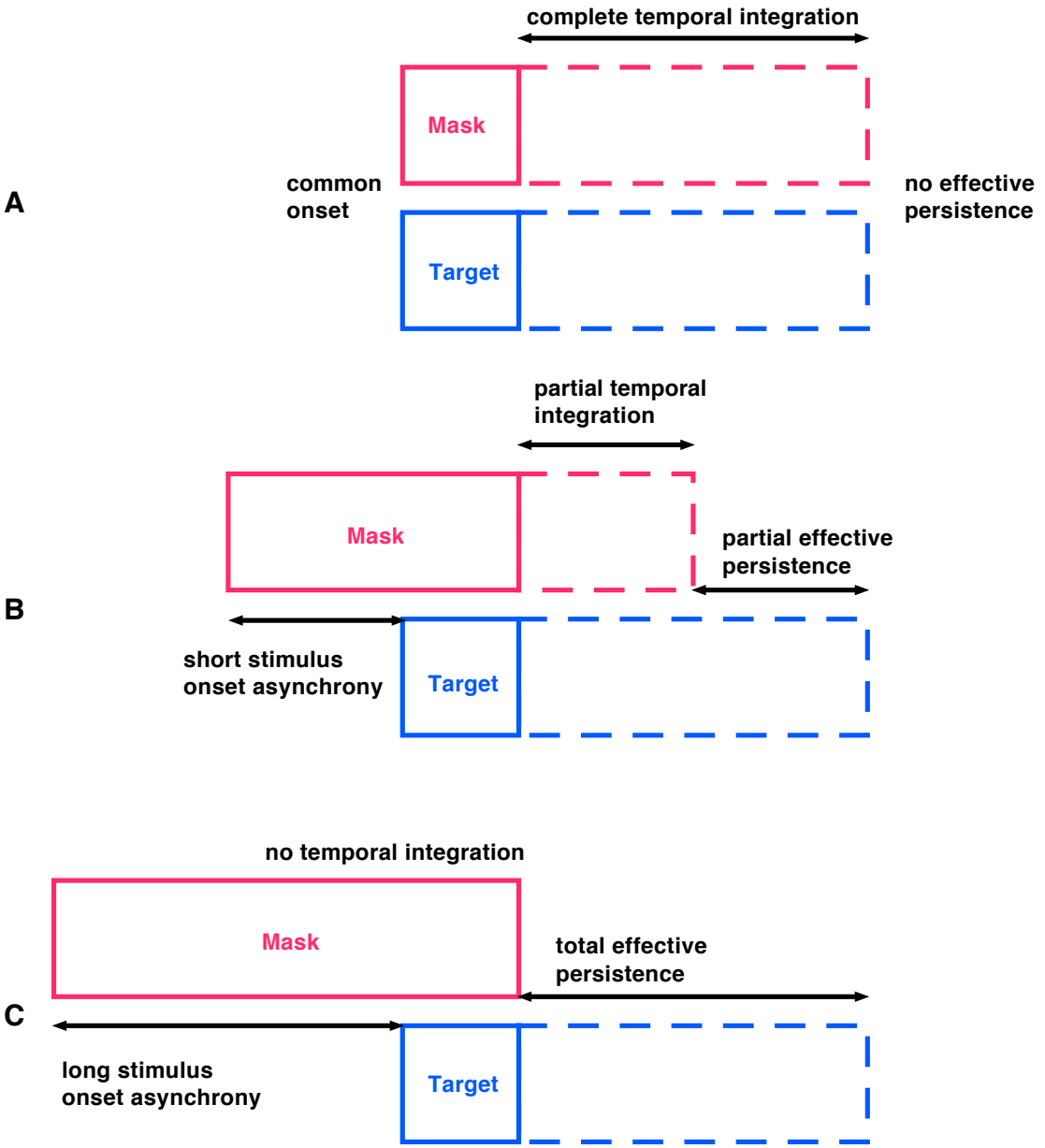


Figure 1.7 Schematic depiction of integration masking

Depending on the stimulus onset asynchrony between mask and target display varying portions of visual persistence of the two stimuli are temporally integrated. **A** Both displays are completely integrated into a single percept since they are presented at the same time. **B** Short stimulus onset asynchronies between mask and target result in partial temporal integration and reduced effective persistence of the target display. **C** When both displays are presented distant in time (> 100-150 ms), temporal integration does not occur and target information is accessible for the full time of its persistence (Figures adapted from Di Lollo, 1980).

Within this temporal window of sensory persistence, however, individual information can also be accumulated. Hence, individuation capacity could be a consequence of the window bandwidth. Fractionating sensory persistence with integration masking provides a means to obtain more fine-grained temporal information about object individuation mechanisms within one perceptual instant.

1.3 Computation, algorithm and implementation of temporal windows in vision

The following three chapters characterize the relationship between temporal windows and informational capacity in visual processing based on Marr's three different, but complementary, levels of description (Marr, 1982).

First, this link is described on the computational level by showing that limiting the time to access the sensory image, by fractionating the perceptual instant, reduces the informational capacity of the observer, unit by unit. Hence, and in contrast to common assumption, I show that reading-out individual and stable object-files is not an instantaneous process. Computationally, the goal of implementing a temporal window might be to accumulate sensory evidence in order to converge within the inherent capacity limitations of the visual percept.

Second, an algorithm is proposed that describes the transformations of the visual signal within the "psychophysical process" (Wundt, 1989) as a succession of temporal epochs from transient sensory detection to the read-out of individual and then identifiable objects. I describe the input-output relationships between three stages in object processing – detection, individuation and identification - and define temporal buffering as the respective boundary condition between stability and flexibility of the perceptual representation (Figure 1.8).

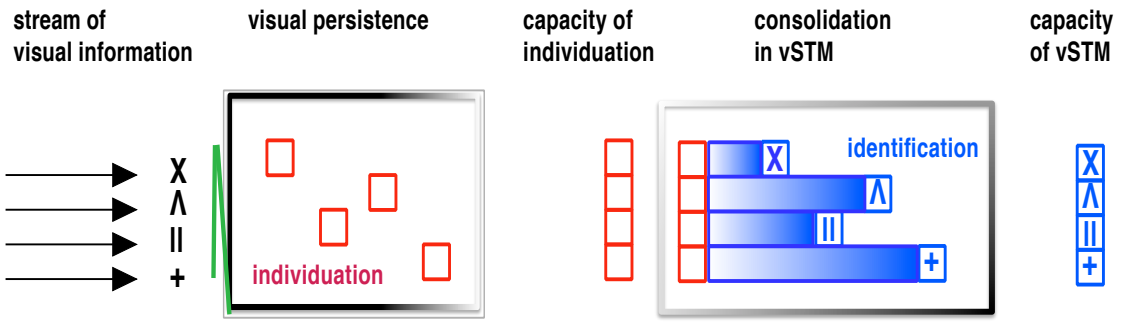


Figure 1.8 Illustration of temporal windows in object processing

Under normal viewing conditions, the stream of visual information is individuated during the period of visual persistence of the sampled sensory image. Items that are individuated are potential "object files" which can then be identified and consolidated into visual short-term memory (vSTM).

Third, the implementation of such a temporal windowing mechanism within neuronal processing is explored as a trace of interactions in sensory persistence within the α-phase locked component of the visual evoked response. I show that phase synchrony mediates between temporal segregation and integration of temporally close visual transients, but this synchrony plays less of a role if the onsets of successive stimuli fall in different temporal integration windows.

Finally, in the last chapter I discuss the findings from the three studies and propose that visual computations within one perceptual instant of time can be described as equilibrating temporal and informational resolution of the environment. Fragmenting the continuous stream of visual information into temporal windows provides a neuronal mechanism to accumulate sensory evidence over time and almost simultaneously read-out spatio-temporally invariant representations. In this way temporal windows provide stability and continuity within a dynamically changing visual environment.

2. Rapid enumeration within a fraction of a single glance:

The role of visible persistence in object individuation capacity

The study reported here has been published under the above title by Wutz, A., Caramazza, A., & Melcher, D. in Visual Cognition (2012).

The number of items that can be individuated at a single glance is limited. Here, we investigate object individuation at a higher temporal resolution, in fractions of a single glance. In two experiments involving object individuation we manipulated the duration of visual persistence of the target items with a forward masking procedure. The number of items as well as their stimulus onset asynchrony (SOA) to the mask was varied independently. The results showed main effects of numerosity and SOA, as well as an interaction. These effects were not caused by a generic reduction of item visibility by the mask. Instead, the SOA manipulation appeared to fractionate the time to access the sensory image. These findings suggest that the capacity limit of 3-4 items found in object individuation is, at least partially, the consequence of the temporal window of access to sensory information.

As noted by Spelke, “the organization of the perceived world into units may be a central task of human systems of thought” (1988, p. 229). Extracting objects from sensory input is called object individuation and involves selecting features from a crowded scene, binding them into a unitary representation and individuating this spatiotemporal unit from other individuals in the image (Treisman & Gelade, 1980; Pylyshyn, 1989; Kahneman, Treisman & Gibbs, 1992; Xu & Chun, 2009). It has long been noted that individuation is limited in capacity: we can quickly and effortlessly perceive that there are exactly three items but not that there are exactly thirty items (Jevons, 1871). Whereas enumeration of five or more items has to rely on serial and time-consuming counting or imprecise estimation, smaller numerosities are presumably simultaneously apprehended by a qualitatively distinct mechanism known as “subitizing” (Kaufmann, Lord, Reese & Volkmann, 1949). There are a number of competing theories for why subitizing, and individuation in general, is limited to sets of only about 3 or 4 (for review, see Piazza, Fumarola, Chinello & Melcher, 2011), but it is widely assumed that this limit arises from a uniform process in one instant.

Here we examine what happens at a smaller time scale, *within* a single glance, to critically evaluate the assumption that subitizing is indeed instantaneous. Previous theories of subitizing have attempted to account for spatial or numerical limits (Pylyshyn & Storm, 1988; Pylyshyn, 1989). We consider an alternative hypothesis based on time, in which the effective duration of the stimulus limits the time available to individuate items. A briefly presented visual display has a limited perceptual persistence during which time it may be processed and categorized (Wundt, 1899; Sperling, 1960; Loftus, Duncan & Gehrig, 1992). Object individuation within a single glance can be viewed, then, as a race for items to be individuated before this window closes. If individuation is time-limited, then a single glance might be too long to reveal the processes underlying

rapid enumeration. We test whether the “magic number” is actually a “magic time period” for the individuation of items being held in a rapidly decaying sensory memory. Our experimental goal was to fractionate time into smaller units to watch the unfolding of the object individuation process.

In order to vary the effective duration of the items on the screen we used a special form of visual masking. Masking can occur when two successive visual stimuli are presented within 100-150 ms from each other, effectively being integrated into a single percept (Di Lollo, 1980; Loftus & Irwin, 1998). Ongoing neural processes still active from the first stimulus can dramatically reduce the visible persistence of the second stimulus, a phenomenon called masking by integration of contours (Di Lollo, 1980). This forward masking manipulation makes it possible to quantitatively change the duration of visual persistence (and iconic memory access) by varying the onset asynchrony between the first and second display. This experimental manipulation provides a means to obtain more fine-grained temporal information about object individuation mechanisms.

To measure object individuation within a fraction of a single glance, we conducted two experiments in which we independently varied stimulus onset asynchrony (SOA) between the forward mask and the target items, as well as the number of items (Figure 2.1). The target items were presented superposed upon the mask and the time course of their visible persistence was altered by the onset asynchrony to the mask display. The first experiment investigated enumeration within and beyond the subitizing range, while the second experiment required participants to identify whether a previously viewed target shape was present among a variable number of shapes. Both experiments shared the selection and individuation of discrete entities from a crowded scene, differing only in later processing stages. As the same masking procedure was used in

both experiments, any similarities in the results between the experiments can be ascribed to the effects of this manipulation on object individuation in isolation from additional mechanisms.

Since enumeration performance is affected by item visibility, with a uniform effect within the subitizing range (Palomares & Egeth, 2010; Palomares, Smith, Pitts & Carter, 2011), we also ran a control condition in order to disentangle the effects of the mask on temporal processing from its more generic effect on item visibility. We report that reducing the effective persistence of the items, unlike other methods that simply reduce item visibility in general, leads to a specific effect within the subitizing range which is consistent with our hypothesis that capacity limits are caused, at least in part, by temporal limits on the individuation process.

2.1 Methods and materials

2.1.1 Subjects

Fourteen observers participated in the enumeration (13 female, mean age *M* = 22.5 y, *SD* = 3.9 y) and the identification experiment on visible persistence (9 female, *M* = 23.9 y, *SD* = 9.1 y). There were 8 participants in the control experiment on item visibility (3 female, *M* = 29.1 y, *SD* = 2.0 y). All participants provided informed consent as approved by the institutional ethics committee, took part in exchange for course credits and had normal or corrected-to-normal vision.

2.1.2 Stimuli and apparatus

The experiments were run on a HP Intel Quad core computer using MATLAB 7.9 (MathWorks, Natick, MA) and Psychophysics Toolbox Version 3 (Brainard, 1997; Pelli,

1997). Participants were seated in a dimly lit room, approximately 50 cm from a 19'' Mitsubishi monitor (1600x1200 resolution) running at 85 Hz. On each trial a different pattern of 200 randomly oriented, partially crossing black lines (luminance: 0.07 cd/m^2; mean line length = 1.2° visual angle, *SD* = 0.24°; mean line width = 0.12°, *SD* = 0.04°; mean size of whole pattern = 18° vertically by 10° horizontally) was presented centered on a white background (luminance: 99.93 cd/m^2; Figure 2.1). This pattern remained on the screen and then after a variable onset delay a variable number of items (1-4 or 6) appeared which were linearly superposed upon the random line pattern by use of the image processing technique 'alpha blending'.

In the experimental trials the random line pattern was always presented with the same alpha-blending values as the target display and with full contrast. There was no contrast difference between mask and target displays; therefore the target display intensity was relatively low. As also the presentation time was quite brief (71 ms), afterimages may have played a negligible role for these kinds of visual stimuli (Di Lollo, 1980). In the enumeration experiment and its control the letter 'X' was used as targets (Figure 2.1 A and 2.1 B). In the identification experiment a variable number of twelve possible two-line drawings (i.e. cross, two parallel lines) was presented, of which one was previously defined as the to be identified shape (Figure 2.1 C). All items were colored in black, were 1.6° of visual angle in height and 1.28° in width and were placed randomly on one of 16 possible locations within an invisible, central rectangle of 8.8° vertical and 7.52° horizontal eccentricity with a minimum buffer of 0.8° between the locations.

2.1.3 Procedure

2.1.3.1 Enumeration with variable mask-item(s) SOA

All subjects received verbal and written instructions about the task and completed fifteen practice trials, in which the random line pattern was made 80% transparent by multiplying its alpha channels by a factor of 0.8. Each trial began with a central fixation dot (black, 0.4°) on a white background for 500 ms, followed by a white blank screen for another 500 ms. Then the random line pattern was presented for one of four durations, in order to control the stimulus onset asynchrony (SOA) between the onset of the mask and the item(s). There were four different stimulus onset asynchronies: 0 ms (common onset), 24, 47 or 141 ms. The target display, in which the item(s) to be enumerated were superposed upon the masking pattern, was always presented for the same brief duration of 71 ms. This target display was immediately followed by a white screen until the participant's response, which was recorded by pressing the corresponding number on a keyboard (Figure 2.1 D).

Although reaction time was recorded, the analyses focused on the proportion of correct trials to avoid any potential effects of participants searching for the correct number key. Our approach is consistent with previous studies, which have measured correct performance while directly manipulating the presentation time of the stimulus, rather than depending on reaction time in order to avoid potential confounds at the response level (Reed, 1973; McElree & Carrasco, 1999). The participants were instructed that one to eight items could be presented, whereas only one to four or six items were actually shown. This manipulation was required in order to prevent a response bias to always report the highest possible numerosity when in doubt, as might have been expected given that the mask contained a large number of elements. Both the behavioral data and an explicit question after the experiment verified that none of the

subjects was aware that there had been no displays with five, seven or eight items. The experiment consisted of eight blocks of 60 trials. Each of the 20 possible combinations of mask-item(s) SOA and target numerosity was shown three times per block in random order. The experiment lasted approximately 45 minutes.

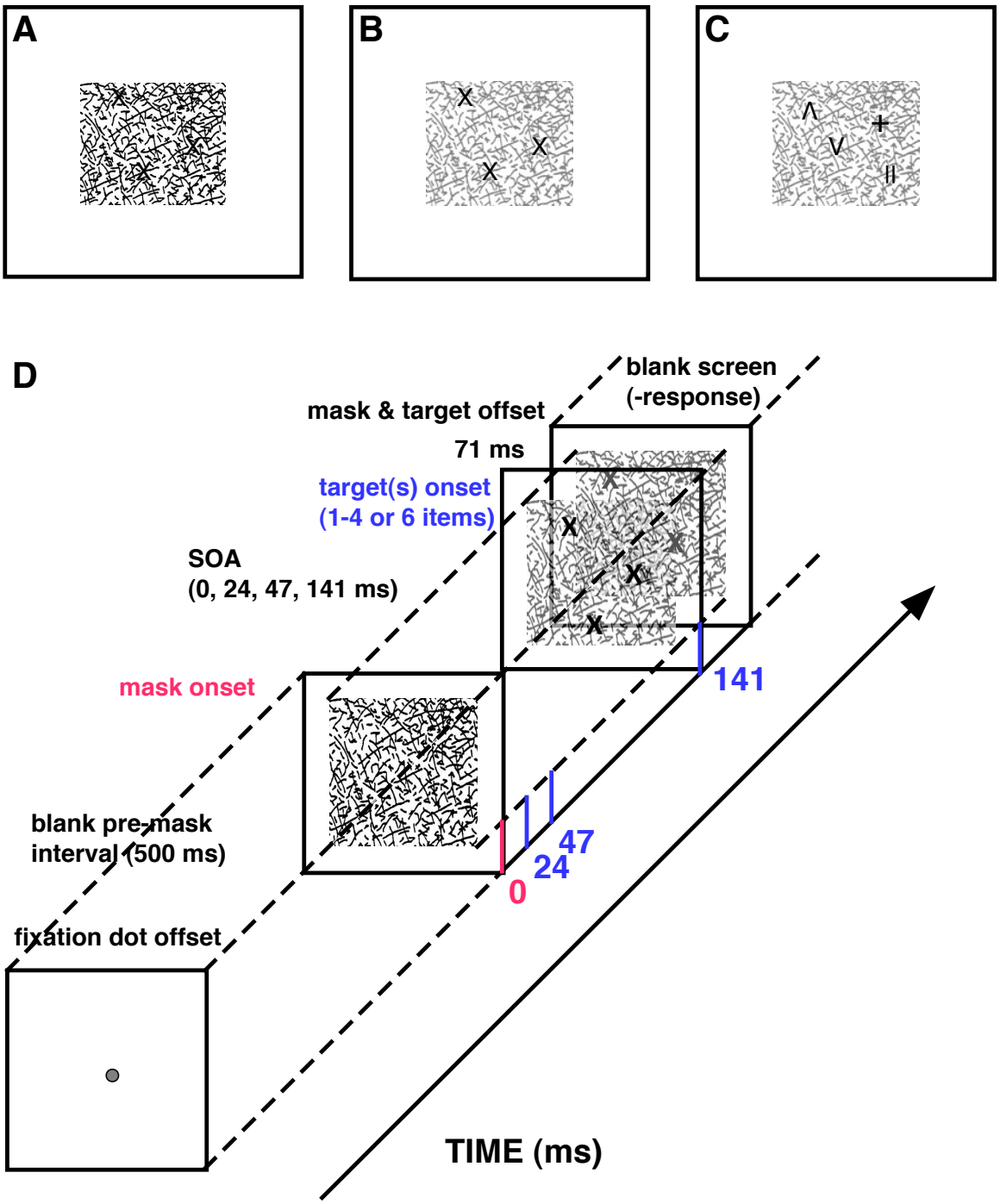


Figure 2.1 Illustration of the mask and target stimuli used in the experiments (panels A-C) and of one trial in the enumeration experiment with the manipulation of SOA (panel D)

A Display with random line pattern and target letters 'X' superposed upon it. Note that the target items are virtually invisible when presented simultaneously with the random line pattern. **B** The same display shown in the left panel but with the random line pattern shown 60% transparent for illustrative reasons. **C** Example of the two-line drawings used as targets, superimposed on the random line patterns shown here at 60% transparence. In the experiments on visible persistence, the random line patterns were always shown at full contrast, as shown in panel A. Mask contrast was an independently varied factor in the control experiment on item visibility. **D** Illustration of one trial in the enumeration experiment with the manipulation of SOA. Throughout the trials, the two independent factors target numerosity (1, 2, 3, 4 or 6) and mask-target(s) SOA (0, 24, 47 or 141 ms) were varied. The targets superposed upon the masking pattern (here shown 60% transparent for illustrative reasons) were always presented for 71 ms, followed by a blank screen until the subject's response.

2.1.3.2 Enumeration with variable mask contrast

A control experiment was conducted to disentangle the impact of the temporal duration of visible persistence from the masking effect on item visibility in general. Instead of varying stimulus onset of the mask relative to the target display, both were presented simultaneously for 71 ms with varying mask contrast. Prior to calculating the contrast values as different proportions along the RGB range, the monitor's luminance in the given settings had been calibrated and gamma corrected. In order to arrive at comparable performance levels between the experiments, the contrast values of the mask were chosen based on pilot studies to be 100%, 40%, 30% and 0% contrast. The condition with 0 ms SOA in the first experiment was identical to the condition with 100% mask contrast in the control experiment (Figure 2.2). Given the five different numerosity levels, there were 20 possible factorial combinations presented within a block. Eight blocks of 60 trials were run. The control experiment lasted around 40 minutes.

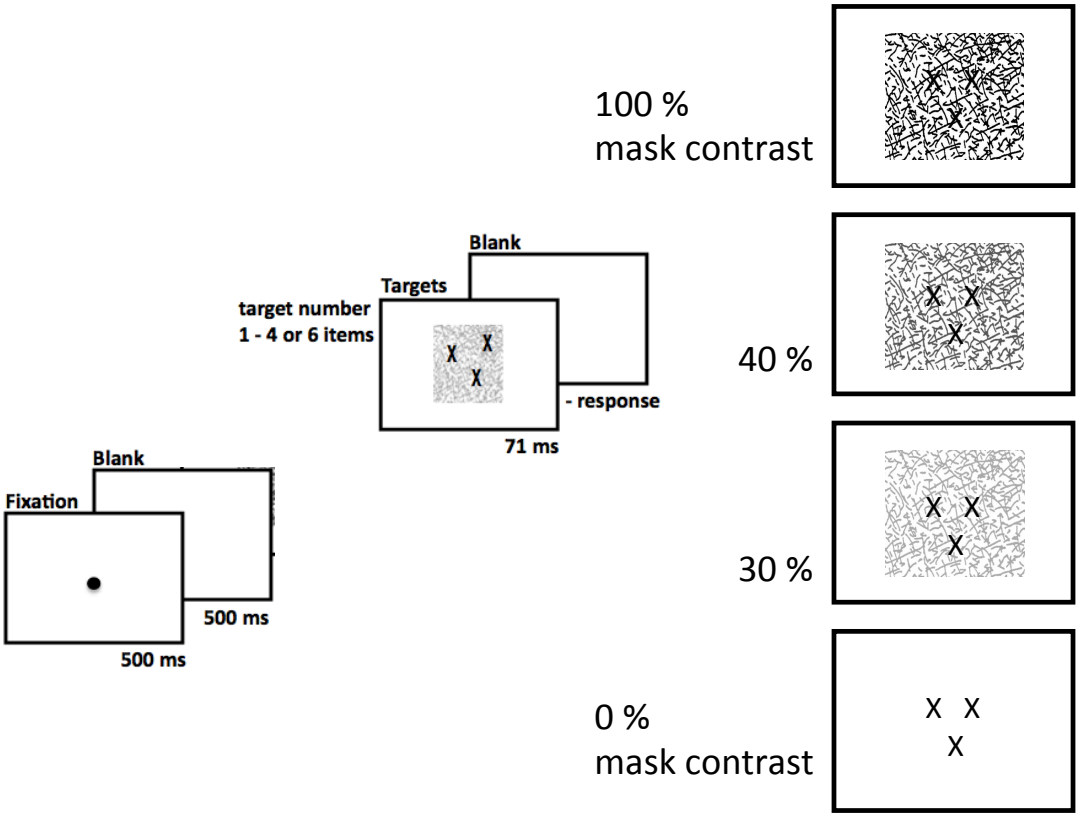


Figure 2.2 Schematic depiction of the enumeration experiment with the manipulation of mask contrast

Throughout the trials, the two independent factors target numerosity (1, 2, 3, 4 or 6) and mask contrast (100%, 40%, 30%, 0%) were varied. The temporal onset of mask and target display was the same and both displays were always presented for 71 ms, followed by a blank screen until the subject's response.

2.1.3.3 Identification with variable mask-item(s) SOA

The procedure for the identification task was the same as in the enumeration experiment except for the following changes. First, a target shape was shown centrally at the beginning of the trial for 500 ms, followed by a 500 ms white blank screen. The task on each trial was to state whether or not the target shape was one of the items presented in the subsequent display. The target shape was present on 50% of the trials. Participants responded by pressing a key corresponding to target absent or present. Based on the

results of the enumeration experiment, and taking account of the additional requirement of identification in this task, the mask-item(s) SOAs were slightly changed with respect to the first experiment to be fit within the range of 24 to 200 ms. Within one block every combination of the three factors -- SOA (24, 47, 71, 200 ms), set size (1-4 or 6) and target presence (present/absent) -- was shown three times and in random order (Figure 2.3). Experiment 2 comprised eight blocks of 120 trials and lasted approximately 90 minutes.

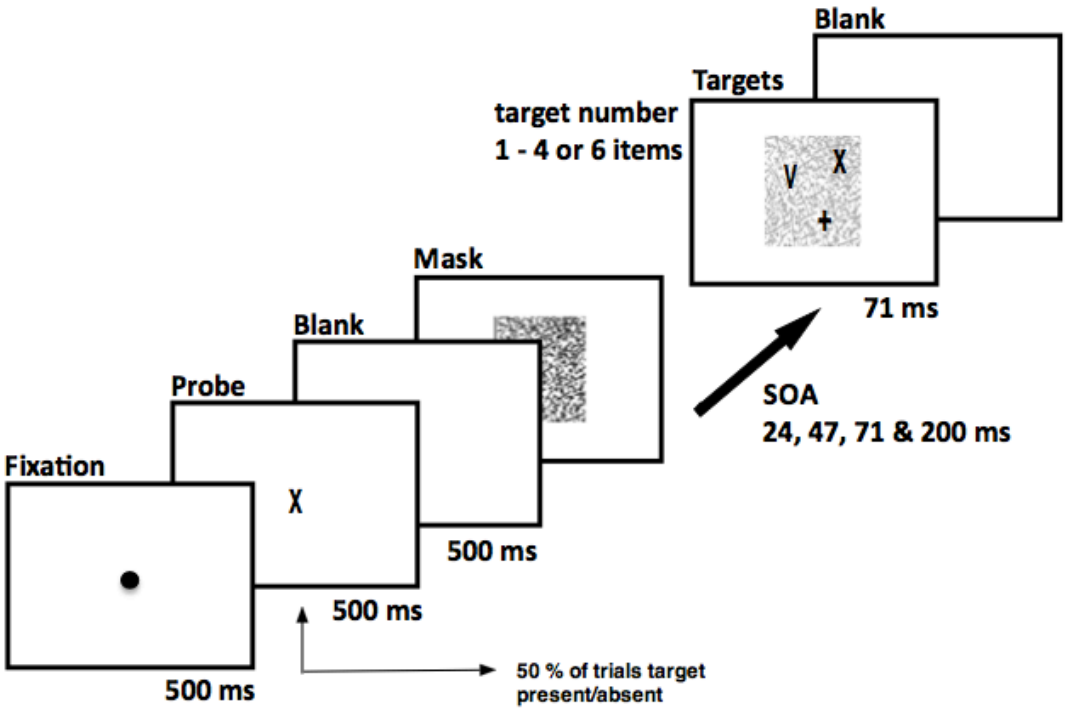


Figure 2.3 Schematic depiction of the identification experiment with the manipulation of SOA

Participants had to identify a previously presented target shape (50 % present/absent) among shapes of different set sizes. Throughout the trials, the two independent factors target set size (1, 2, 3, 4 or 6) and mask-target(s) SOA (24, 47, 71 or 200 ms) were varied. The targets superposed upon the masking pattern (here shown 60% transparent for illustrative reasons) were always presented for 71 ms, followed by a blank screen until the subject's response.

2.1.4 Data analysis

As the study was designed to investigate object individuation within the subitizing range, data for numerosities from one to four items were fed into a two-way (masking level - by - number) within-subject analysis of variance (ANOVA) for all reported experiments. The residuals of all reported variables were normally distributed as shown by a Kolmogorov-Smirnov test. In case sphericity for the given factor was not tenable, F-Ratios have been adjusted with a Greenhouse-Geisser correction. To further investigate interactions between the two factors, post-hoc t-tests between performance at each numerosity (1-4) and a baseline condition (see below) were conducted (p-values Bonferroni-corrected). Due to technical difficulties, reaction time data for the enumeration experiment with variable mask-item(s) SOA were available for only twelve of the fourteen subjects.

2.2 Results

2.2.1 Enumeration with variable mask-item(s) SOA

In contrast to previous studies showing good enumeration performance up to about 4 items, the masking manipulation used here led to a dramatic effect on proportion correct (Pc) and reaction times (RT) even within the subitizing range (Figure 2.4 A & B, supplemental Figure SF2.1). This effect is confirmed by a within-subjects ANOVA on the accuracy and the reaction times, which revealed main effects of SOA (Pc: $F(3, 39) = 198.9$, $p < .001$, $\eta_p^2 = .939$; RT: $F(1.7, 18.7) = 33.7$, $p < .001$, $\eta_p^2 = .754$) and item numerosity (Pc: $F(3, 39) = 14.5$, $p < .001$, $\eta_p^2 = .526$; RT: $F(1.6, 17.9) = 29.9$, $p < .001$, $\eta_p^2 = .731$), as well as an ordinal interaction between these two factors (Pc: $F(9, 117) = 5.1$, $p < .001$, $\eta_p^2 = .283$; RT: $F(3, 33.5) = 7.4$, $p < .001$, $\eta_p^2 = .402$). As expected, enumeration accuracy increased and reaction times generally decreased for smaller item numerosities and longer SOA. As subitizing capacity can vary between three and

four items across participants, a similar ANOVA with a subitizing range of up to three items was calculated and comparable results were obtained (see supplemental Table ST2.1).

Visual inspection of Figure 2.4 A & B confirms a qualitative difference in performance between small and large numerosities (Kaufman et al., 1949; Piazza et al., 2011). Accuracy meliorated less for six items (39.6 % increase) compared to four (64.5%) and reaction times for six items increased with longer SOAs. However, the different SOA conditions affected enumeration differently even for small numerosities within the subitizing range. To better understand these differences, average performance for one to four items at each SOA was used as a baseline condition (BL) for subsequent paired comparisons. The mean proportion of correct trials across numerosities is the expected value given stochastic independence of the probability of a correct response and the specific number of items within the subitizing range (for a definition of stochastic independence see Pearson, 1900). In other words, within the subitizing range the probability of a correct response should not depend upon the specific number of items shown on the screen—indeed, the equality of accuracy and RTs within the subitizing range has been the defining aspect of the concept of subitizing. Any deviations from this value of stochastic independence between numerosity and response are indicative of an effect of SOA.

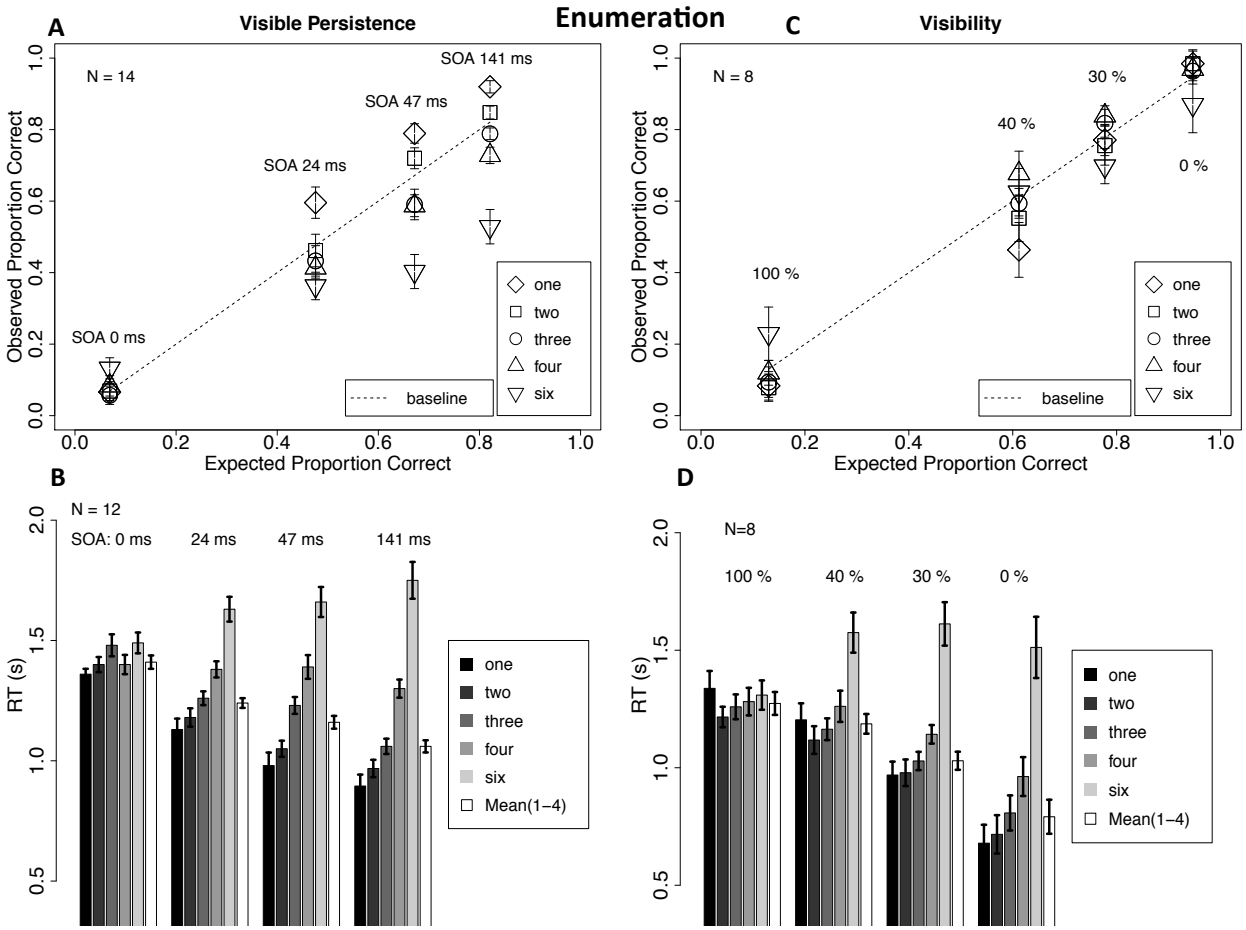


Figure 2.4 Results of the enumeration experiments with the manipulation of SOA (panels A & B) and mask contrast (panels C & D)

A Observed proportion as a function of expected proportion of correct trials, given stochastic independence of the probability of a correct response and numerosity within the subitizing range at every level of mask-Item(s) SOA for different item numerosities. **B** Reaction times at each mask-item(s) SOA for different item numerosities and for the average reaction time within numerositiy one to four. **C** Observed proportion as a function of expected proportion of correct trials, given stochastic independence of the probability of a correct response and numerosity within the subitizing range at every level of mask contrast for different item numerosities. **D** Reaction times at each mask contrast for different item numerosities and for the average reaction time within numerositiy one to four. Vertical deviations from the dashed lines indicate differences between observed and expected values. Error bars display one standard error of the mean for within-subject designs. Individual performance values have been centered on the mean performance of each subject before calculating the standard error.

The forward mask was effective at limiting the effective duration of the target stimulus. In the case of 0 ms SOA, enumeration accuracy did not exceed expected performance at chance level (12.5 % correct) for all item numerosities (*min* = 6 %, *max* = 13 % correct; one-tailed $t(13)$s < 0.3, ds < 0.09) and reaction times were generally quite high. Behavioral performance therefore indicates a high level of uncertainty within the observers, confirming the phenomenological experience that targets were virtually invisible when the mask and the item(s) were presented simultaneously.

As SOA increased to 24 ms, accuracy improved for all item numerosities within the subitzing range (M = 47.6 %), but most strongly for one-item displays (59.6 %). Only accuracy for one item was higher than the baseline ($t(13) = 4.2$, $p < .005$, $d = 1.1$), whereas the other numerosity levels (two, three and four items) showed no significant difference (abs ($t(13)$s) < 1.9, abs (d)s < 0.5). Reaction times for one item were significantly lower than the baseline ($t(11) = -3.1$, $p < .05$, $d = -0.9$). Overall, these results show that the small increase in SOA affected object individuation most strongly for one-item displays. In other words, at the 24 ms SOA the assumption of stochastic independence of target set size, within the subitizing range, was violated.

For a mask-item(s) SOA of 47 ms one-item and, marginally significant, two-item displays (showing a 26 % increase compared to the 24 ms SOA) were more accurately enumerated than the baseline (two vs. BL: $t(13) = 2.6$, $p < .095$, $d = 0.7$). Reaction time data revealed the same pattern of results: One and two items yielded faster reaction times compared to the baseline average (both $t(11)$s < -4.3, $p < .005$, both ds < -1.2). These results suggest that there was a particular benefit in enumeration for one and two item displays with the 47 ms SOA condition.

At the longest SOA tested, performance for three item displays finally approached the baseline level. Accuracy and reaction times for four item displays were

still significantly worse than the baseline (Pc: $t(13) = -4.9$, $p < .001$, $d = -1.3$; RT: $t(11) = 5.5$, $p < .001$, $d = 1.6$). Therefore, at the 141 ms SOA there was an additional improvement in performance for the three-item displays.

2.2.2 The role of temporal effects of masking versus a general reduction in visibility

As described above (see Introduction and Methods), a control condition varying mask contrast was used to distinguish between time constraints on enumeration and a more general effect of reduced visibility. The pattern of results (Figure 2.4 C & D, supplemental Figure SF2.2) shows that reducing item visibility *per se* had a quite different effect on enumeration compared to those reported above with variable mask-item(s) SOA. One particularly obvious difference between the two conditions is shown by greater accuracy for large numerosities (6 items) in the contrast control task when the mask contrast was high. Good performance for six items reflects a bias towards reporting higher responses under this condition, perhaps due to confusing the mask with the target. This finding is interesting because it shows that better performance for small numerosities in the main experiment, described above, was not due to a tendency to guess a small number when visibility was poor. When the mask was not presented at all (0 % contrast), accuracy was equally high for all set sizes within the subitizing range. In addition, RTs showed a clear qualitative distinction between small and large numerosities, even though the slope was not completely flat within the subitizing range (Figure 2.4 C & D). Therefore the observed enumeration performance with these stimuli in this unmasked condition fits well into the existing literature (see Folk, Egeth, & Kwak, 1988; Mandler & Shebo, 1982).

Reducing mask contrast from 100 to 30 % led to an increase in enumeration accuracy ($F(1.2, 8.6) = 89.2$, $p < .001$, $\eta_p^2 = .927$) and decrease in reaction times ($F(2,$

14) = 9.0, $p < .005$, $\eta_p^2 = .564$) within the subitizing range (Figure 2.4 C & D). Furthermore, for both accuracy and reaction times, a main effect of numerosity was observable (Pc: $F(3, 21) = 4.5$, $p < .02$, $\eta_p^2 = .393$; RT: $F(3, 21) = 4.3$, $p < .02$, $\eta_p^2 = .381$). As the pattern of this effect, however, is quite the opposite for these two measures (Figure 2.4 C & D), enumeration performance cannot really be distinguished within the subitizing range with respect to a possible speed-accuracy trade-off. Most importantly, the two factors (mask contrast and item numerosity) within the subitizing range, did not interact (Pc: $F(6, 42) = 1.9$, $p > .1$, $\eta_p^2 = .210$; RT: $F(2.5, 17.3) = 1.4$, $p > .2$, $\eta_p^2 = .166$; Figure 2.4 C & D). Thus, the overall trend showing that manipulating item visibility in general had a uniform effect across small item numerosities was consistent with previous studies (Palomares & Egeth, 2010; Palomares et al., 2011). These results suggest that the effect of masking on enumeration observed in the first experiment is not simply due to alterations in item visibility in general but to constraints on the temporal aspects of visual processing, namely the time course of visible persistence of the to be enumerated items.

2.2.3 Identification

In the first experiment, object individuation was operationalized by enumeration. Of course, enumeration is a complex task. Therefore, it was useful to include a second task, which shared the first two stages of processing (selection and individuation) with Experiment 1 but differed in later stages. Thus, the second experiment isolated individuation from the “numerical cognition” aspects of enumeration and added an additional identification component.

Despite the difference in tasks, the overall trend was remarkably similar. Both reaction times and the proportion of correct trials (which includes hits and correct

rejections) were significantly altered by mask-item(s) SOA (Pc: $F(3, 39) = 47.7$, $p < .001$, $\eta_p^2 = .786$; RT: $F(1.6, 21.2) = 14.2$, $p < .001$, $\eta_p^2 = .521$) and set size (Pc: $F(3, 39) = 30.7$, $p < .001$, $\eta_p^2 = .702$; RT: $F(3,39) = 29.0$, $p < .001$, $\eta_p^2 = .690$). An interaction was found between SOA and set size for the accuracy measure (Pc: $F(9, 117) = 3.6$, $p < .002$, $\eta_p^2 = .215$; Figure 2.5). Again, post-hoc t-tests between each numerosity from one to four and their mean at every level of SOA were conducted to highlight the pattern of interactions of the masking manipulation within the subitizing range (Figure 2.5, supplemental Figure SF2.3).

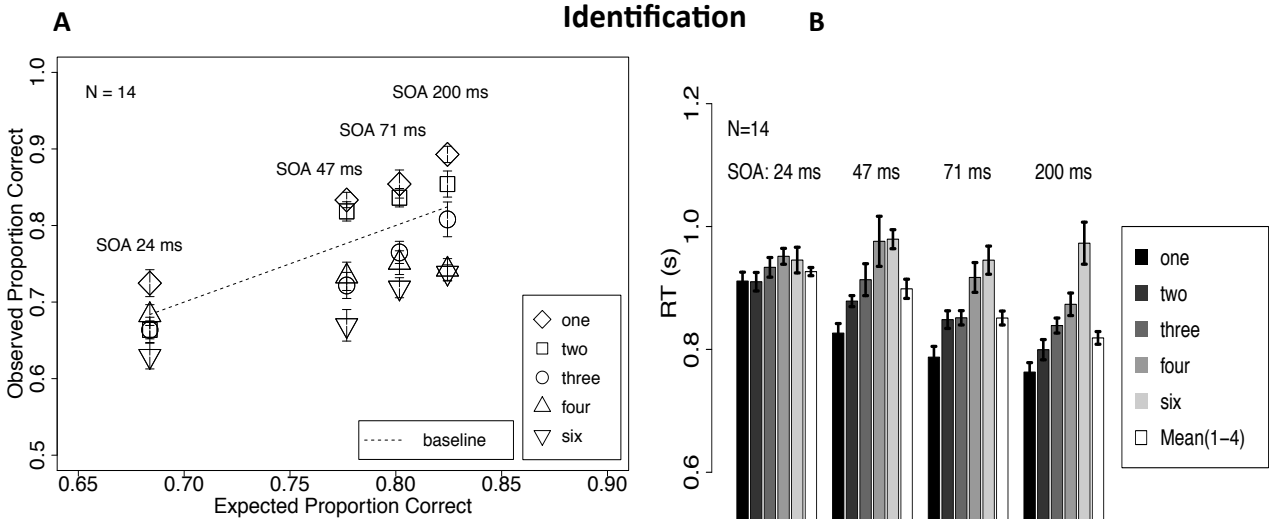


Figure 2.5 Results of the identification experiment with the manipulation of SOA

A Observed proportion as a function of expected proportion of correct trials, given stochastic independence of the probability of a correct response and set size within the subitizing range at every level of mask-Item(s) SOA for different item set sizes. **B** Reaction times at each mask-item(s) SOA for different item set sizes and for the average reaction time within set sizes one to four. Vertical deviations from the dashed line indicate differences between observed and expected values. Error bars display one standard error of the mean for within-subject designs. Individual performance values have been centered on the mean performance of each subject before calculating the standard error.

Accuracy for a single item was higher than baseline performance at the 24 ms SOA ($t(13) = 3.3$, $p < .025$, $d = 0.9$), but this was found only for the single item condition (abs($t(13)$s) < 1.9, abs (d)s < 0.5). This confirms the particular benefit in the individuation of one item with a very short SOA found in the enumeration task (Experiment 1).

With 47 ms and 71 ms SOA, accuracy for one- and two-item displays were significantly above the baseline average (all $t(13)$s > 2.9, $p < .05$, all ds > 0.75). Reaction times for one item were faster than the baseline for both SOAs (both $t(13)$s < -3.8, $p < .01$, both ds < -1.0). The striking difference in identification accuracy for two-item displays compared to baseline performance suggests that, as in the enumeration experiment, there was a shift in the number of items preferentially processed.

Accuracy and reaction times for three-item displays converged towards baseline performance only at the longest SOA (200 ms). Identification for four items remained less accurate (74.3 %; $t(13) = -5.7$, $p < .001$, $d = -1.5$) and slower (0.87 s; $t(13) = 5.3$, $p < .001$, $d = 1.4$) than baseline. As the performance measures for larger set sizes (4 & 6) seem to saturate, this pattern of results suggests an increase in capacity as a function of SOA with a limit of around three items. It is important to note that the persistent one item benefit in enumeration was also found with a binary (present/absent) response. This finding strongly argues against the possibility that a tendency to report the ordinal extremes (1 or 8 items) completely explains the results of the first experiment (see also supplemental Figure SF2.4 for response matrices at each SOA). Moreover, a table showing reaction times and accuracy for each experiment is included in the supplementary materials (ST 2.2 – 4).

2.3 Discussion

The main finding was that the masking procedure affected performance within the subitizing range. This effect was observable in two tasks that both required object individuation but differed in response selection. Thus, it is likely that masking interacted with the individuation of multiple objects and not with subsequent response-limited processes. Furthermore, this effect was not caused by generic alterations in item visibility, as shown by the control condition of the first experiment. Instead, the manipulation of the SOA appeared to temporally fractionate the effective persistence of the visual image and this limited the capacity of object individuation. Thus, theories that try to explain the "magical number four" by a limit in a simultaneous process may under-sample its timescale. We suggest that a more thorough analysis of the temporal dynamics of individuation might help to explain capacity limitations.

It is important to note that the effect of mask-target(s) SOA cannot be explained by an improvement in the visual system's readiness to process temporally trailing displays. Di Lollo (1980) showed that presenting a mask with a variable SOA, but changing also the mask configuration simultaneously with target display onset, disrupted performance regardless of SOA. Based on that earlier result we can exclude attentional pre-cueing as a major determinant of the current pattern of findings.

Although we focus here on rapid individuation, rather than memory, for objects, previous studies of visual working memory have also reported an effect of time (Gegenfurtner & Sperling, 1993; Vogel, Woodman & Luck, 2006). In those earlier studies, a backward mask was used to limit the display duration of multiple items in a visual working memory paradigm. Our study differs in several ways. First, we focus on rapid individuation, rather than consolidation of already individuated items into memory. Our task does not require subjects to remember the identity of multiple items, only the

numerosity of multiple items or the identity of a single item. Second, we examined the first tens of milliseconds of visual processing, while the earlier studies—which were interested in higher cognitive aspects of working memory—focused on mental processes happening after 100 ms. In other words, the earlier studies investigated what happened after the glance, while we explored the unfolding of object individuation within the glance.

Here, we examined object individuation as a process in which multiple objects race to emerge from a complex scene as unique, individual objects, within a very short window of time. However, in interpreting a time-limited process it is important to acknowledge the mathematical complexity of determining whether unobservable processes are parallel or serial based on input-output relationships or its statistics (Townsend, 1971). The finding that subitizing is not purely instantaneous, but *evolves within a single glance*, could be accounted for by either a serial or parallel mechanism. Vanishing access to sensory information could limit the time to serially repeat a number of actions. A theoretical implementation of such a serial mechanism to extract information from a visual scene was proposed by Ullman (1984). Elemental operations, like shifting the processing focus or indexing a salient item, are combined into sequences or *visual routines* to allow real-time execution of computationally complex tasks, like enumeration. The subitizing phenomenon therefore may reflect the cardinal number of such a *visual routine* applied upon the sensory image during the time of its persistence.

On the other hand, the duration of sensory memory could constrain a parallel process that converges into a correct percept above a specific intensity threshold: “the greater the number of objects to which our consciousness is simultaneously extended, the smaller is the intensity with which it is able to consider each” (Hamilton, 1859, p.

164). It is therefore reasonable that the time required for a temporally evolving, parallel process to reach threshold depends on the number of items it processes. Processing intensity in visual neurons can be modulated by attention (Moran & Desimone, 1985). Given that subitizing has been demonstrated to require attention (Egeth, Leonard & Palomares, 2008; Olivers & Watson, 2008), the computational speed of object individuation in parallel may be a function of the degree to which attentional resources have to be shared among multiple items. Increasing attentional load e.g. by increasing target-distractor similarity (Watson, Maylor, Allen & Bruce, 2007) or adding an attention demanding dual task (Olivers & Watson, 2008), might slow down the core individuation process to an extent that the read-out of information cannot be accomplished within the time in which the sensory input is available to the mechanism at work. In these situations, one might expect that the observers rely on counting or estimation mechanisms even for small numbers of items, instead of specialized subitizing, which is in general accordance with recent findings (Burr, Turi & Anobile, 2010). Our results are consistent with, but go beyond, recent evidence for a role of attention in subitizing by providing testable hypotheses for how and when attention might limit subitizing performance.

In a similar way, competitive interactions between potential proto-objects in a type of saliency map could explain numerosity-dependent processing rates for a parallel mechanism. When there is only one salient object, the proto-object would emerge *in a fraction of a single glance.* Competition among multiple items may require more time to converge into a stable percept both at the stage of individuation and at the level of memory (Dempere-Marco, Melcher & Deco, 2011). Subitizing, therefore, might be explained if the duration of the decay of sensory information was on average equal to the time necessary to process four items in parallel.

If sensory input is available indefinitely to the observer, e.g. under unlimited viewing conditions, a new cycle of read-out of information can be initiated after the initial glance, in order to refresh the initial sensory image. One example is “counting”, a process which is generally considered to require multiple perceptual steps and the use of saccadic eye movements (Kowler & Steinman, 1977). When the sensory image contains more informational units than those individuated during the “initial glance”, an increase in both reaction times and eye movement frequencies for item numerosities above the subitizing range would be expected (Watson, Maylor & Bruce, 2007).

Independent of the processing mechanism - serial or parallel - the present results show that object individuation is not a temporally uniform process across the subitizing range. We suggest that capacity limits in individuation are caused, at least in part, by temporal constraints on the underlying mechanism. The rate of temporal processing for individuation would likely depend on the stimuli used and on the individual subject. The analysis of the temporal dynamics of object individuation evolving in *fractions of a single glance* might therefore lead to an explanation of subitizing as revealing a “magical time period”, rather than a “magical number”.

3. Temporal buffering and visual capacity:

The time course of object formation underlies capacity limits in visual cognition

The study reported here has been published under the above title by Wutz, A., & Melcher, D. in Attention, Perception, & Psychophysics (2013).

Capacity limits are a hallmark of visual cognition. The upper boundary in our ability to individuate and remember objects is well known, but - despite its central role in visual information processing - not well understood. Here we investigate the role of temporal limits in the perceptual processes of forming ‘object files’. Specifically, we examined the two fundamental mechanisms of object file formation - individuation and identification - by selectively interfering with visual processing using forward and backward masking with variable stimulus onset asynchronies (SOAs). While target detection was almost unaffected by these two types of masking, they showed distinct effects on the two different stages of object formation. Forward ‘integration’ masking selectively impaired object individuation, whereas backward ‘interruption’ masking only affected identification and the consolidation of information into visual working memory. We therefore conclude that the inherent temporal dynamics of visual information processing are an essential component in creating the capacity limits in object individuation and visual working memory.

One of the fundamental goals of perception is to enable us to interact with objects in the environment. According to Wundt, the interaction of an observer with the external environment (the "psychophysical process") can be subdivided into three temporally successive and distinct stages (Wundt, 1899, 1900). The first stage ("perception") describes the entrance of an object into the field of vision, allowing it to be detected. In a subsequent stage, termed "apperception", the addressed object occupies the focus of the observer's attention. Finally, the observer develops the volition to react to the object either cognitively, by storing it into memory, or behaviorally with a grasping or a saccadic eye movement.

Wundt's description emphasizes how object recognition involves a temporal succession of distinct processing stages - from an unlimited in capacity, but fragile, purely bottom-up and in parallel computed sensory representation (iconic memory: Sperling, 1960, 1963; Neisser, 1967) to a capacity limited, durable and cognitively structured visual store (visual short-term memory: Sperling, 1960, 1963; Phillips & Baddeley, 1971) leading to a an action that results in an isomorphic one-to-one relation between observer and object.

As shown in Figure 3.1 A, Wundt's stage of apperception can be further subdivided into two processing mechanisms: object individuation and object identification (Xu & Chun, 2009). Individuation involves selecting features from a crowded scene, binding them into a unitary representation and individuating this spatiotemporal unit from other individuals in the image (Treisman & Gelade, 1980; Pylyshyn, 1989; Kahneman, Treisman & Gibbs, 1992; Xu & Chun, 2009). Object representations at this stage are suggested to be coarse and contain only minimal feature information (Xu & Chun, 2009). Some of these 'object files' (Kahneman et al., 1992) are elaborated subsequently during object identification. It is at this stage that

identity information becomes available to the observer and the content of the object files can be consolidated into durable and reportable representations in visual working memory. The number of objects available at this stage is variable, depending on object complexity, task demand and representation resolution (Alvarez & Cavanagh, 2004; Xu & Chun, 2009). As individuation precedes identification, the capacity of the latter has its upper bound in the limit of the former (Melcher & Piazza, 2011; Piazza, Fumarola, Chinello & Melcher, 2011, Dempere-Marco, Melcher & Deco, 2012).

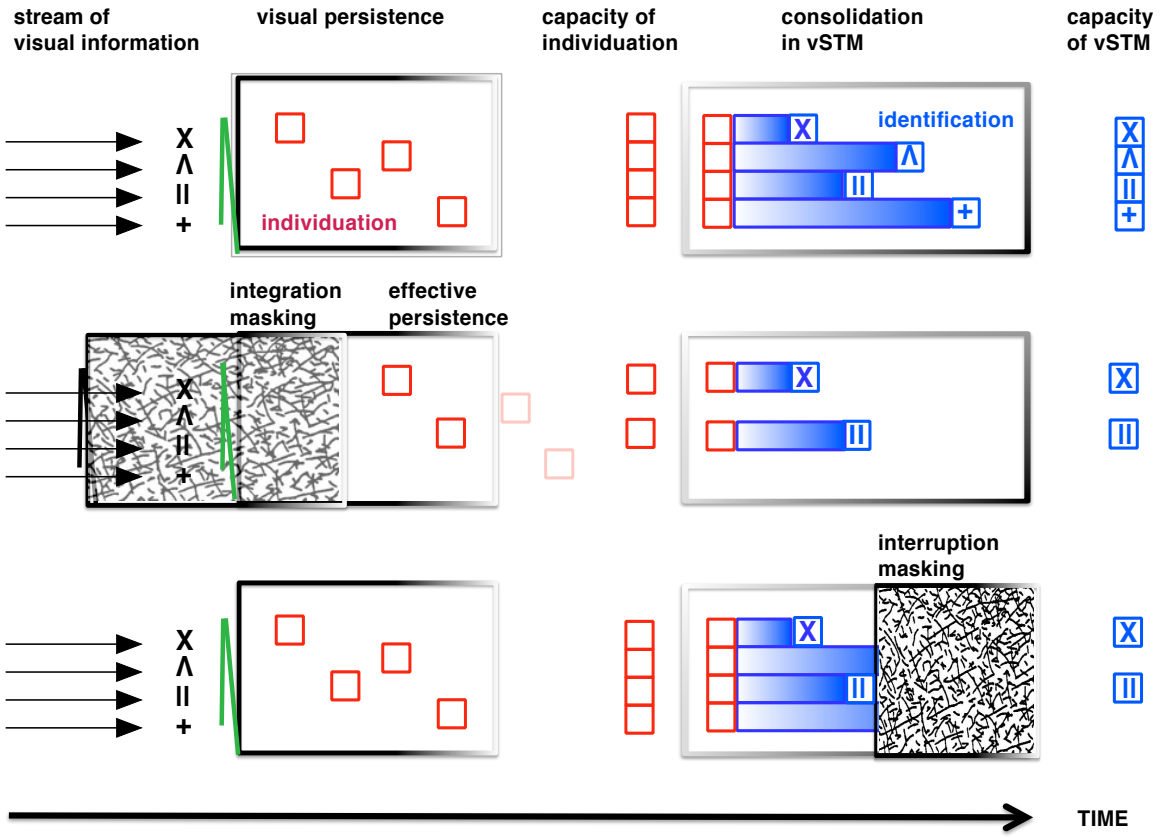


Figure 3.1 Illustration of how the temporal limits of visual object processing can result in capacity limits for individuation and identification

A Under normal viewing conditions, the stream of visual information is individuated during the period of visual persistence of the sampled sensory image. Items that are individuated are potential "object files" which can then be identified and consolidated into visual short-term memory (vSTM). **B** Integration masking via forward masking reduces the effective persistence of the target items, leading to a reduction in capacity for individuation and consequently also for identification. **C** Interruption masking does not

influence the initial individuation of items but instead disrupts the identification and consolidation of items into vSTM.

The goal of the present report is to investigate whether capacity limitations in object processing can be traced to temporal constraints on the distinct object processing stages. We therefore try to embed the ongoing debate about the roots of capacity limits in vision (reflected in the 'subitizing' phenomenon (Jevons, 1871; Kaufman et al, (1949)) and visual working memory (Luck & Vogel, 1994; Cowan, 2000) into the already well-established body of work about the temporal dynamics of the visual system (Wundt, 1989; Sperling, 1960, 1963; Loftus, Duncan & Gehrig, 1992).

Specifically, we use two types of masking—integration and interruption masking—in order to influence either the individuation or identification stage of object file formation. Visual masking refers to the reduction of the visibility of one stimulus, called the target, by another stimulus shown before and/or after it, called the mask (Enns & Di Lollo, 2000; Breitmeyer & Öğmen, 2006). It is usually explained in terms of a two-factor theory: integration and interruption masking (Scheerer, 1973a, b). Integration masking occurs when target and mask information are combined together, as a consequence of the imprecise temporal resolution of the visual system. Integration masking can occur with either forward or backward masking for short SOA values (up to around 100 ms between the target and the mask). In contrast, 'interruption masking' affects higher- level mechanisms that are engaged in object recognition and yields a J-shaped masking function, as it can only occur for masks appearing temporally after the target display (Enns & Di Lollo, 2000; Breitmeyer & Öğmen, 2006). The effect of this kind of masking is thought to reflect a disruption of processing *after* perceptual analysis is already completed, but *before* the representation has been consolidated into visual working memory (Vogel, Woodman & Luck, 2006).

Our hypothesis is that integration masking should selectively affect the individuation stage by reducing the effective persistence of the target items (Figure 3.1 B). Integration masking is very effectively implemented with a specific forward masking technique that makes it possible to quantitatively change the duration of visual persistence (and iconic memory access) and the degree of temporal integration by varying the onset asynchrony between the first and second display (Di Lollo, 1980). Also, in the case of backward masking with a very short SOA we would expect integration masking to occur and to limit the effective visual persistence of the target and thus the individuation processes.

In contrast, we predict that interruption masking should selectively affect the identification of items, after individuation has largely finished, since the consolidation of targets into vSTM would be interrupted (Figure 3.1 C). Interruption masking should only occur for backward masking with longer SOAs (greater than around 100 ms). We therefore expect to see a specific influence of such backward masking on visual memory but not on individuation.

We investigated the two stages of object file formation (individuation and identification/consolidation) using the two forms of contour masking (integration and interruption) in a fully counter-balanced two-by-two design. In order to watch the temporal unfolding of object file formation, we employed forward and backward masking techniques, using a variety of SOAs, in two tasks: enumeration and change detection. Enumeration serves as an operationalization of object individuation, whereas change detection serves as the main paradigm for studying visual working memory.

If capacity limits in vision and visual working memory can be explained by temporal constraints on the formation of object files, we would expect that techniques that limit processing time at specific temporal stages of the visual analysis would

selectively inhibit the successive mechanisms operating upon the sensory input at these stages. In other words, integration masking should selectively impair object individuation, whereas interruption masking should only affect object identification and the consolidation of object information into visual working memory. This design allows us to test the role of temporal dynamics in the individuation and identification of objects.

We also included a control condition to measure the effects of the forward and backward masking paradigms on a simple detection task. This control condition was necessary to ensure that reduced performance from masking reflected not simply the fact that the targets were effectively invisible, but reveals limits on visual computations *within* the sensory image aimed to arrive at a structured, object-like representation. This control condition allowed us to study the unfolding of object representations, from simple detection of the presence of a stimulus, to the individuation of a specific number of target items and then eventually the recognition of object file content.

3.1 Methods and materials

3.1.1 Subjects

Sixteen participants (11 female, mean age *M* = 22.9 y, *SD* = 4.2 y) completed the series of four conditions in the main experiment on object file formation. A different group of ten subjects (6 female, mean age *M* = 22.7 y, *SD* = 3.7 y) participated in the control condition measuring target detection. All participants provided informed consent, as approved by the institutional ethics committee. Subjects took part in exchange for course credit or a small payment and had normal or corrected-to-normal vision.

3.1.2 Stimuli and apparatus

The experiment was run on a HP Intel Quad core computer using MATLAB 7.9 (MathWorks, Natick, MA) and Psychophysics Toolbox Version 3 (Brainard, 1997; Pelli, 1997). Participants were seated in a dimly lit room, approximately 45 cm from a 19'' Mitsubishi monitor (1600x1200 resolution) running at 85 Hz. On each trial a different pattern of 400 randomly oriented, partially crossing black lines (mean line length = 1° visual angle, mean line width = 0.1°, mean size of whole pattern = 13.4°) was presented centered on a white background (Figure 3.2 A). In the forward masking conditions this pattern remained on the screen and then after a variable onset delay a variable number of items (up to 6) appeared which were linearly superposed upon the random line pattern by use of the image processing technique 'alpha blending' (Figure 3.2 B). The physical properties of both mask and target elements, i.e. contrast, mean line length, mean line width, were equated. Furthermore the 'alpha blending' procedure edited the transparency/opacity values of the visual stimuli assuring a mathematically correct superimposition of local element contrast, without creating any discontinuities in luminance which would have been a cue to finding the target. Mask and target elements differed only in their temporal onset in order to exclusively vary the amount of integration masking. Thus, this method combined both forward and simultaneous masking. In the backward masking conditions the same random line pattern was presented with a variable ISI with respect to target offset (Figure 3.2 A). The same set of 12 possible two-line drawings (i.e. cross, two parallel lines) was used as items in all four experimental manipulations and also in the control conditions (Figure 3.2 B & C). All items were colored in black, were 0.9 ° of visual angle in size and were placed randomly on one of 16 possible locations within an invisible, central rectangle of 5.4° of visual angle in eccentricity with a minimum buffer of 0.6° between the locations.

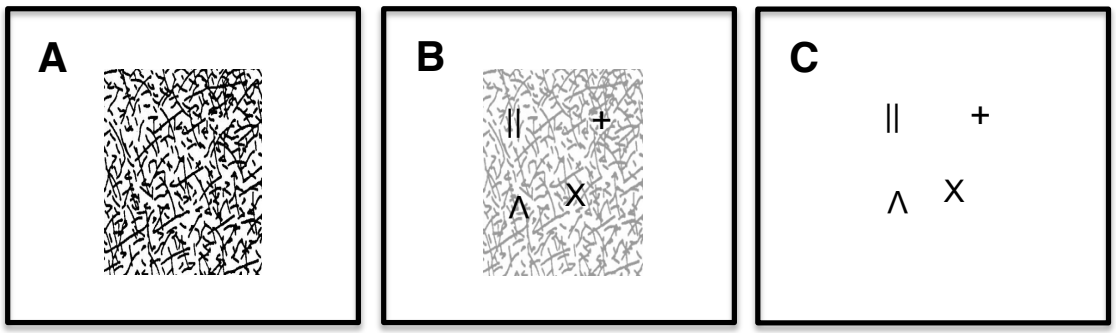


Figure 3.2 Illustration of the mask and target stimuli used in the experiment (panels A-C) **A** Display with random line pattern used as masks in the experiment. **B** Example of the two-line drawings used as targets, superimposed on the random line pattern shown here at 60% transparence for illustrative reasons. In the experiment the random line pattern was always shown at full contrast, as shown in panel A. **C** Example of the two-line drawings used as targets upon a blank white screen like in the backward masking conditions.

3.1.3 Procedure

Each subject completed the four experimental manipulations in two sessions consisting of two conditions each and comprising approximately 1.5 hours each. The serial order of the four different experimental manipulations (masking technique (forward vs. backward) crossed with task (enumeration vs. change detection)) was fully balanced across the observers in a Latin square design (Figure 3.3). Groups of four subjects completed one of the four counterbalanced sequences within the Latin square. Prior to the experiment the full set of possible target items was presented to the subjects on the screen for an unrestricted viewing time. All subjects received verbal and written instructions about each task and completed twenty practice trials for each condition. In all four conditions, each trial began with a central fixation dot (black, 0.3°) on a white background for 500 ms, followed by a blank white screen for another 500 ms. Then, the order of events in the trial depended upon the masking technique and task, as explained below. The subject's response on the keyboard initiated the next trial.

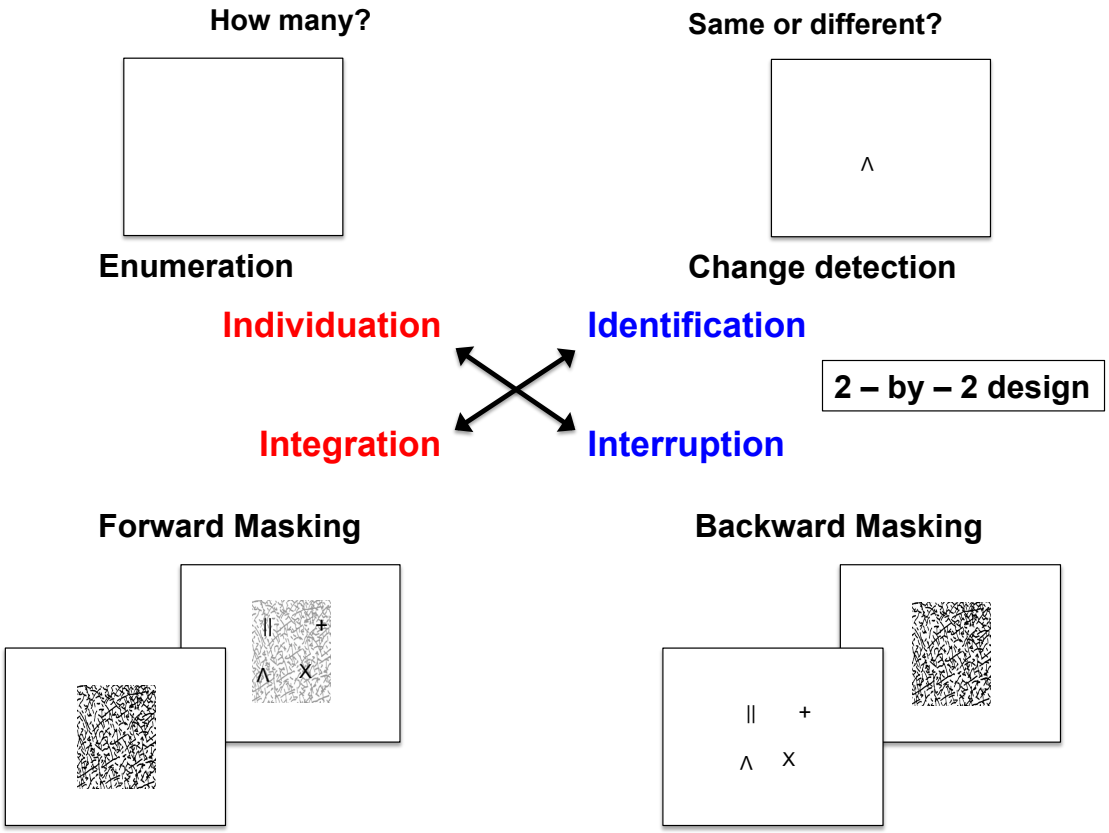


Figure 3.3 Schematic depiction of the 2 – by – 2 design employed in the experiment

Each of the four conditions (forward masking with enumeration, forward masking with change detection, backward masking with enumeration, backward masking with change detection) was administered to the participants using a fully counterbalanced Latin square design. Enumeration served as an operationalization of object individuation, while change detection measured object identification. The technique of forward masking is considered to favor integration masking (Di Lollo, 1980) and backward masking with longer SOAs has an interrupting influence on visual performance (Scheerer, 1973a, b).

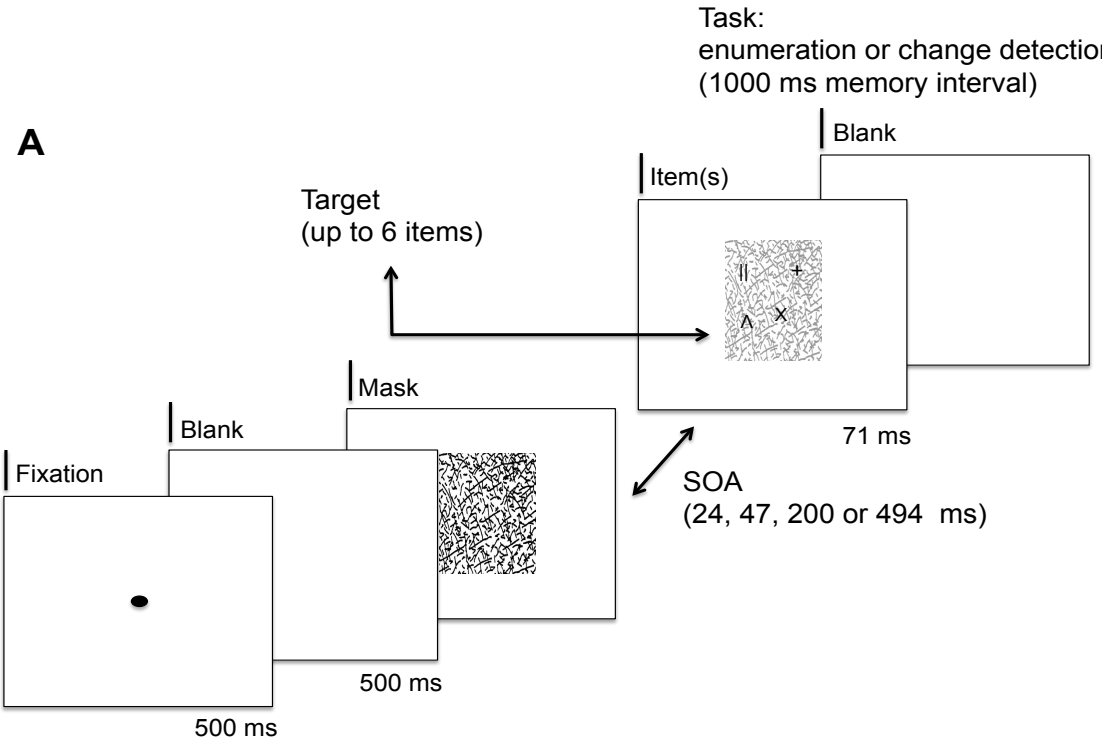
3.1.3.1 Forward masking vs. backward masking

In the case of forward masking, the random line pattern was presented for one of four durations, in order to control the stimulus onset asynchrony (SOA) between the onset of the mask and the item(s). There were four different stimulus onset asynchronies: 24, 47, 200 or 494 ms. The target display with the items to be enumerated or memorized was

superposed upon the masking pattern and was always presented for the same brief duration of 71 ms (Figure 3.2 B). The target display was immediately followed by a white screen (Figure 3.4 A). Using this procedure we achieved an optimal temporal resolution of the visual mechanisms operating within the first tens of milliseconds around target exposure during which integration masking mostly occurs, as very short SOAs can be used. This simultaneous mask makes it possible to fractionate the time course of visible persistence of the target items. It is important to note that rather than merely reducing item visibility; the combination of forward and simultaneous masking specifically affects the rate at which objects are individuated. At short SOAs, only one object can be individuated, while with increasing SOAs, object capacity increases in steps (see Chapter 2; Wutz et al., 2012).

On backward masking trials the target items were shown first for 71 ms upon a white background, followed by the random-line pattern after a variable SOA. Unlike the forward masking technique, the target and the masking displays were not presented simultaneously (Figure 3.2 C). Four different SOAs were used: 71 ms (immediately after target offset), 118, 200 or 506 ms. Any delay period between target offset and mask onset was filled by the presentation of a blank white screen. The mask was always shown for 71 ms and immediately followed by a white screen (Figure 3.4 B). The 71 ms SOA mask condition was included in order to fit within the temporal limits of integration masking, while the longer SOAs were expected to result in interruption masking.

Forward Masking



Backward Masking

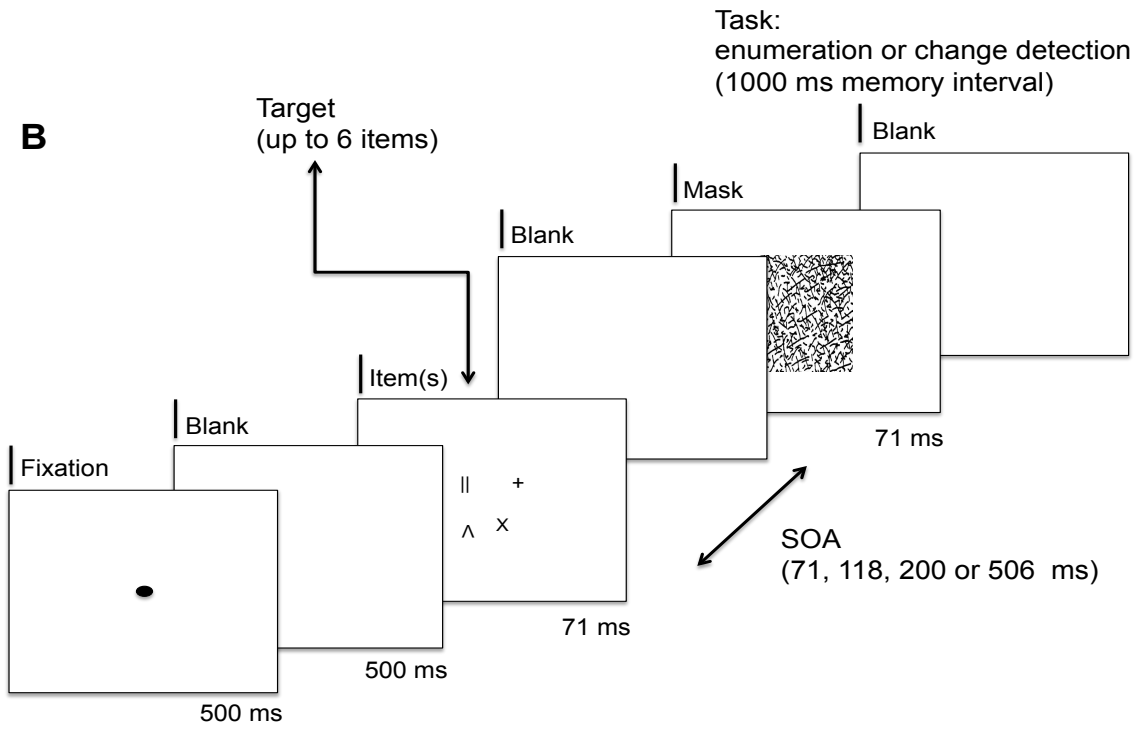


Figure 3.4 One trial for forward (panel A) and backward masking conditions (panel B)

A Illustration of one trial in the forward masking condition. Throughout the trials, the two independent factors target set size (1, 2, 3, 4 or 6 for enumeration; 2, 4 or 6 for change detection) and mask-target(s) SOA (24, 200 or 494 ms) were varied. The targets superposed upon the masking pattern (here shown 60% transparent for illustrative reasons) were always presented for 71 ms, followed by a blank screen until the subject's response in the enumeration condition. With change detection a memory interval of 1000 ms followed the target display, followed by a probe item for 71 ms. **B** Illustration of one trial in the backward masking condition. Throughout the trials, the two independent factors target set size (1, 2, 3, 4 or 6 for enumeration; 2, 4 or 6 for change detection) and mask-target(s) SOA (71, 118, 200 or 506 ms) were varied. The targets were always presented for 71 ms, followed in case of SOAs bigger than 71 ms by a blank screen and a mask for 71 ms. In the enumeration condition a blank screen followed until the subject's response. With change detection a memory interval of 1000 ms followed the target display, followed by a probe item for 71 ms.

3.1.3.2 Enumeration vs. change detection

Both masking techniques were used in a crossed design with two different task demands: enumeration or change detection within the item display. In the case of enumeration, the subjects had to indicate the number of perceived items by pressing the corresponding number on a keyboard immediately after target or mask offset. Whereas one to four or six items were actually shown (there were never five targets), the participants were instructed to respond within the full range between one and six items. We did not inform participants that there were no trials with five target items in order to avoid a guessing strategy in which participants would always respond "six" when the number of items exceeded their subitizing range. The enumeration condition consisted of 6 blocks of 60 trials. Each of the 20 possible combinations of SOA and target numerosity was shown three times per block in random order.

On change detection trials a probe was presented, after a blank delay of one second, for 71 ms in one of the locations that were previously occupied by a target item. This memory interval of one second was always held constant regardless of the

temporal position of the mask and the item display. The identity of the probe matched the corresponding item in the target set in 50% of the trials. Participants responded by pressing a key corresponding to probe identity being the same or different. Note the difference between the identification condition in the previous chapter and the change detection task here (both unfortunately termed “identification”). In chapter 2, the probe item was shown before the onset of the target display, which required the observers to select and individuate multiple items and only match the single probe to the target set. Here the entire set of items has to be individuated, identified and remembered until memory probe onset. These task demands go beyond initial individuation and require the consolidation of identity information from multiple items into vSTM. Within one block every combination of the three factors - SOA, set size and probe identity - was shown three times and in random order. The conditions using change detection comprised five blocks of 72 trials.

3.1.3.3 Target detection

In order to clearly disentangle the effects of masking on the formation on object-files from a more generic effect on target display visibility, we ran a control condition requiring the participants to simply detect the target display. Participants reported whether or not at least one target had been presented on each trial. Each of the ten subjects was run in this control task under forward and backward masking conditions in a single session. The order of masking type was balanced across subjects. All participants received verbal and written instructions about the task and completed one practice block for each condition. The trial sequence and the masking procedures used in this control condition were identical to those described above, except for the following changes: Only two SOAs were used, the shortest and longest ones described in the experimental

procedures above. This means that for forward masking there were SOAs of 24 and 494 ms, while for backward masking we used SOAs of 71 (immediately after target offset) and 506 ms. Target displays were presented on 50% of the trials. Within these target-present trials, the display consisted with equal probability of either one or four targets presented for 71 ms. On the other half of the trials (target-absent trials), the target display was replaced either with an instance of the masking pattern (on forward masking trials) or with a white screen (on backward masking trials) for an equal duration (71 ms). The subjects were instructed to press a previously specified key indicating the presence or absence of a target display, irrespective of the number of targets, after mask or target offset, respectively. Within one block every combination of the two factors – SOA and target presence/absence – was shown sixteen times and in random order. For both forward and backward masking three blocks of 64 trials each were run. The whole session comprised approximately 45 minutes.

3.1.4 Data analysis

For all experimental conditions the proportions of correct trials were fed into a two-way within-subjects ANOVA with the factors set size (1-4 & 6 for enumeration; 2,4 & 6 for change detection) and stimulus onset asynchrony (SOA). In case the residuals of one variable within one condition did not follow a normal distribution as indexed by a Kolmogorov-Smirnov test, the analysis for this condition was repeated using a Friedman test. As the main results did not differ between parametric and non-parametric procedures, only the ANOVA results are reported. If sphericity for a given factor was not tenable, F-Ratios have been adjusted with a Greenhouse-Geisser correction. The alpha-level for post-hoc planned comparisons has been corrected with a Bonferroni procedure.

For better comparability of the results between the different conditions involving

object file formation, the proportions of correct trials with set size four have been translated into corresponding capacity estimates for each SOA. This calculation was based upon performance measures for four-item displays in accordance with previous reports (e.g. Vogel, Woodman, & Luck, 2006) since visual object capacity is likely to converge towards asymptote for this set size (Cowan, 2000). The computation of the capacity estimates takes into account the different guessing rates within the different response measures used (enumeration and change detection). For change detection capacity *K* has been calculated using the following formula:

$$K = (H+CR-1)*N$$

where *K* indicates capacity, *H* hit rate, *CR* correct rejection rate and *N* the number of items in the display (Cowan, 2000).

For enumeration, a guessing correction for a 6-alternative forced choice procedure was applied on the raw proportion of correct trials (Klein, 2001). Capacity estimates were then derived from multiplying these values with the number of items in the display, as explicit in the following formula:

$$K = ((P_{cor}-1/M)/(1-1/M))*N$$

, with *K* capacity, P_{cor} proportion of correct trials, *M* number of alternatives (here: 6), *N* number of items in the display.

3.2 Results

3.2.1 Visual masking and object file formation

In all four conditions of the main experiment there was a main effect of both set size and SOA on the proportion of correct responses (Table 3.1). These main effects confirm the evident trend, in Figure 3.5, of improved performance for longer SOAs and for smaller

set sizes. There was a significant interaction in three of the four conditions (Table 3.1). The ordinal order of the main effects, however, is preserved despite these interactions (Figure 3.5).

Table 3.1 Results of the two-way within-subjects ANOVA for the four crossed conditions in the main experiment: forward masking with enumeration, forward masking with change detection, backward masking with enumeration and backward masking with change detection.

	Factor	*df1*	*df2*	*F*	$p <$	η_p^2
Forward Masking - Enumeration	SOA	3	45	94.912	.001	.864
	Set Size	1.749	26.230	49.475	.001	.767
	SOA x Set Size	5.576	83.643	1.710	.135	.102
Forward Masking - Change Detection	SOA	3	45	9.222	.001	.381
	Set Size	2	30	63.330	.001	.809
	SOA x Set Size	6	90	2.392	.035	.138
Backward Masking - Enumeration	SOA	1.529	22.935	56.031	.001	.789
	Set Size	1.530	22.958	64.574	.001	.811
	SOA x Set Size	3.801	57.020	12.490	.001	.454
Backward Masking - Change Detection	SOA	3	45	34.722	.001	.698
	Set Size	2	30	130.376	.001	.897
	SOA x Set Size	6	90	2.784	.017	.157

For each of the condition's main and interaction effects the degrees of freedom of the numerator, the degrees of freedom of the denominator, the F value, the significance

level and the goodness of fit of the general linear model are displayed.

Increasing the SOA between the forward mask and the to be enumerated items altered performance within the subitizing range (mean(1-4 items), SOA 200 vs. 24 ms: $t(15) = 9.939$, $p < .001$). For all set sizes performance reached a plateau by around 200 ms (mean(1-4 items), SOA 494 vs. 200 ms: $t(15) = 1.161$, not significant (n.s.); Figure 3.5 A). For change detection this amelioration of performance with increasing SOA was only observable for two-item-displays and continued up to 494 ms of SOA (two items, SOA 494 vs. 24 ms: $t(15) = 5.338$, $p < .001$). Visual working memory for higher set sizes did not benefit extraordinarily from an increased SOA (mean(4, 6 items), SOA 494 vs. 24 ms: $t(15) = 2.207$, n.s.; Figure 3.5 B). This pattern of results suggests that the forward masking procedure successively affected the individuation of multiple items, eventually limiting the consolidation of information into visual working memory in a very early level of visual processing.

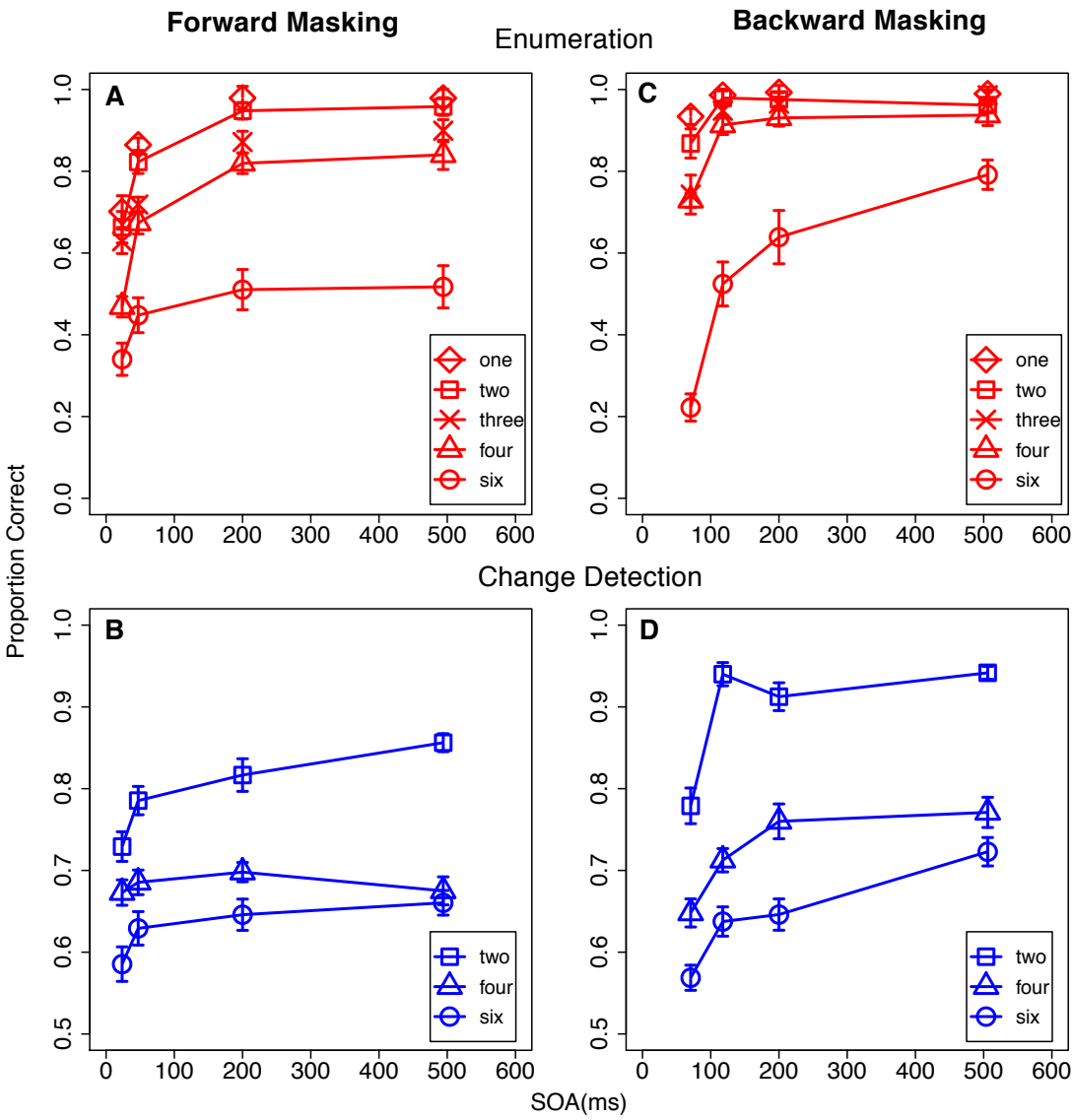


Figure 3.5 Results of the four crossed conditions in the main experiment: forward masking with enumeration (panel A), forward masking with change detection (panel B), backward masking with enumeration (panel C) and backward masking with change detection (panel D)

A Proportion of correct trials as a function of forward mask SOA for different item set sizes for enumeration. For reasons of better comparability to the change detection condition only performance values for the set sizes 2, 4 and 6 are connected with solid lines across different SOAs. **B** Proportion of correct trials as a function of forward mask SOA for different item set sizes for change detection. **C** Proportion of correct trials as a function of backward mask SOA for different item set sizes for enumeration. For reasons of better comparability to the change detection condition only performance values for the set sizes 2, 4 and 6 are connected with solid lines across different SOAs. **D** Proportion of correct trials as a function of backward mask SOA for different item set sizes for change

detection. Error bars display one standard error of the mean for within-subject designs. Individual performance values have been centered on the mean performance of each subject before calculating the standard error.

Results with forward and backward masking differed in two main ways. First, the forward masking conditions had generally lower performance, perhaps due to the effect of the simultaneous mask. This simultaneous mask allows us to study the time course of individuation by creating a limit on the degree to which features can be extracted for multiple objects simultaneously.

Second, the backward mask effects were most noticeable with larger set sizes (six items for enumeration or four items for change detection. Consistent with our hypothesis, this is particularly true within the time course of interruption masking (SOA > 100 ms), Within the subitizing range, increasing SOA from 118 to 506 ms did not improve enumeration performance (mean(1-4 items), SOA 506 vs. 118 ms: $t(15) = 1.464$, n.s). However, for the larger set size (6 items), there was a significant improvement for longer SOAs (six items, SOA 506 vs. 118 ms: $t(15) = 5.374$, $p < .001$; Figure 3.5 C). Similarly, in the case of change detection, there was no benefit from larger SOAs for two item displays (two items, SOA 506 vs. 118 ms: $t(15) = 0.151$, n.s.) while for four and six item displays performance was better with the longest SOA (mean(4, 6 items), SOA 506 vs. 118 ms: $t(15) = 4.829$, $p < .001$; Figure 3.5 D). Given the fact that backward masking had an effect at longer SOAs and larger set sizes, this is consistent with previous suggestions of a specific effect on the consolidation of object file content (Gegenfurtner & Sperling, 1993; Vogel, Woodman & Luck, 2006).

For backward masking, only masks presented immediately after target offset (71 ms SOA), within the range of integration masking, influenced enumeration within the subitizing range (mean(1-4 items), SOA 118 vs. 71 ms: $t(15) = 4.900$, $p < .001$). For

longer SOAs, however, enumeration performance was already at ceiling (see above for the non-significant effect of mean(1-4 items), SOA 506 vs. 118 ms; Figure 3.5 C). Together with the results of the forward masking, this pattern of results is consistent with the idea that subitizing is not instantaneous but rather depends on the effective duration of the stimulus (Wutz et al., 2012). In a similar way, change detection performance for two-item-displays was only altered by this very short SOA and reached asymptote thereafter (two items, SOA 118 vs. 71 ms: $t(15) = 6.203$, $p < .001$; see above for the non-significant effect of two items, SOA 506 vs. 118 ms). Thus, these results suggest that object identification can occur to a limited extent temporally in parallel with or very fast after individuation. The typical four-item limit in visual short-term memory (Luck & Vogel, 1994; Cowan, 2000), however, is not reached within this very short period of time. Visual working memory measures for higher set sizes increased gradually with increasing backward mask SOA (see above for the significant effect of SOA 506 vs. 118 ms; Figure 3.5 D).

3.2.2 Visual masking and target detection

Neither the forward nor backward mask showed the same dramatic reduction in performance for detection as had been found in the main experiment with individuation or identification. For both forward and backward masking, detection performance was above 90 % for almost all set sizes and SOAs both in case of correct rejections in target-absent trials (set size 0) and hits in target-present trials (Figure 3.6). However, for both forward and backward masking a significant effect of SOA was observable (forward masking: $F(1,9) = 17.778$, $p < .002$, $\eta_p^2 = .664$; backward masking: $F(1,9) = 10.494$, $p < .01$, $\eta_p^2 = .538$). A major component of these effects is due to worse performance for one-item displays with short SOAs (forward masking: one item, long vs. short SOA: $t(9)$

= 3.017, $p < .03$; backward masking: one item, long vs. short SOA: $t(9) = 3.074$, $p < .026$; Figure 3.6). This pattern of results resembles enumeration performance under the influence of masking (Figure 3.5 A). For one-item displays detection conceivably is also the main component for enumeration. Therefore it is reasonable that these two conceptually very similar conditions yield comparable results under forward and backward masking. In other words, detection is a limiting factor in the enumeration of one-item displays.

In general, however, the average d' were high under all conditions (forward masking & short SOA: M=3.588, SD=1.266; forward masking & long SOA: M=4.930, SD=0.924; backward masking & short SOA: M=4.347, SD=1.037; backward masking & long SOA: M=5.487, SD=1.013). It is important to note that both forms of visual masking – forward and backward – yielded similar results: target detection was not greatly affected by these masking techniques. This strikingly good detection performance contrasts with the significant masking effects with enumeration and change detection, even though the same temporal parameters in terms of SOA and visual stimuli were used. These results are consistent with the control experiment reported in our recent study of rapid individuation, which also showed that forward and simultaneous masking did not simply reduce target visibility indiscriminately (Wutz et al., 2012).

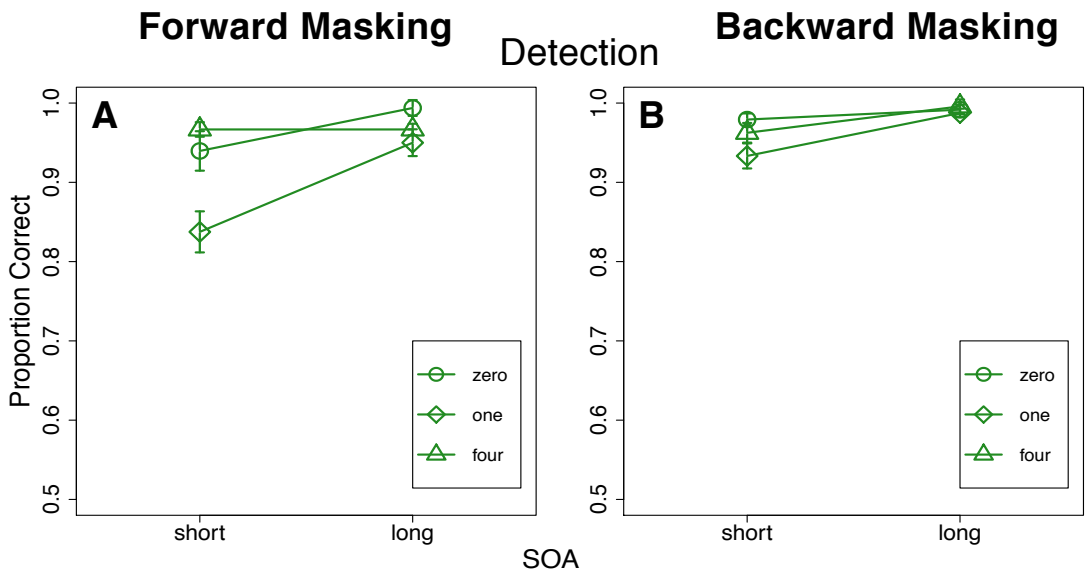


Figure 3.6 Results of the control conditions: forward masking with detection (panel A) & backward masking with detection (panel B)

A Proportion of correct trials as a function of forward mask SOA for different item set sizes for detection. **B** Proportion of correct trials as a function of backward mask SOA for different item set sizes for detection. Error bars display one standard error of the mean for within-subject designs. Individual performance values have been centered on the mean performance of each subject before calculating the standard error.

A second critical difference between the results of the control study and those of the main experiment is that the worst performance is found with one-target displays under forward masks with short SOAs, compared to performance with four items. This is the opposite trend from the enumeration conditions in the main experiment, in which performance was better for one item than four. Displays with one item were harder to detect than those with higher set sizes (SOA 24 ms, one vs. four items: $t(9) = -5.127$, $p < .001$), whereas in the main experiment smaller set sizes were easier to enumerate compared to higher numerosities (SOA 24 ms, one vs. four items: $t(15) = 4.751$, $p < .001$). As target detection either was not affected at all by masking or showed the reverse pattern of results compared to enumeration, the powerful effects of visual

masking on the formation of object files reported above cannot be explained by a failure to register the presence of a target display. Instead, the reported results reveal distinct effects of integration and interruption masking on the extraction of object-like representations from the sensory signal after it has already been registered by the observer as new input, reflecting temporal limits on the perceptual computations *within* the sensory image for the time of its persistence.

3.2.3 Visual masking and object capacity

In order to better understand the accumulation of object information over time, within and beyond the period of visual persistence, we compared object capacity estimates (see Methods) across the four conditions (Figure 3.3 & 3.7). Consistent with a recent study, capacity limits were higher for the enumeration task compared to the visual working memory task (Piazza et al., 2011). Of particular interest, however, are the temporal dynamics of these capacity differences, showing a clear dissociation between forward/integration and backward/interruption masking in the two tasks. Whereas enumeration capacity increased throughout the whole time course of the forward masking procedure, backward masking influenced enumeration only at the very short SOA immediately after target offset (in the time period of integration masking). Visual working memory capacity, however, did not increase as a function of forward mask SOA (staying flat at around 1.5 items), but rose gradually with a longer SOA to the backward mask up to more than 2 items (Figure 3.7).

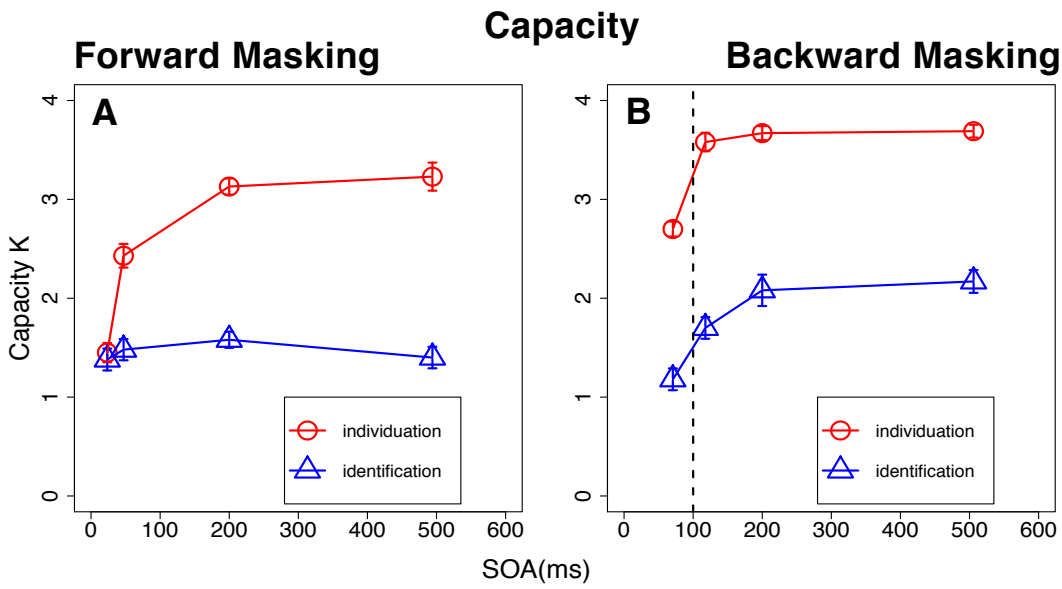


Figure 3.7 Results of the four crossed conditions in the main experiment translated into capacity K for forward masking (panel A) and backward masking (panel B)

A Capacity K as a function of forward mask SOA for enumeration/individuation (red) and change detection/identification (blue). **B** Capacity K as a function of backward mask SOA for enumeration/individuation (red) and change detection/identification (blue). The vertical dashed line indicates the temporal window of visible persistence (Di Lollo, 1980) where the influence of backward masking switches from integration to interruption masking (Scheerer, 1973a, b). Error bars display one standard error of the mean for within-subject designs. Individual performance values have been centered on the mean performance of each subject before calculating the standard error.

This reasoning is confirmed by a within-subjects ANOVA on the capacity estimates for the two tasks within the respective time course of integration and interruption masking. The applied forward masking technique was specifically designed to vary integration masking. For backward masking, however, a distinction between short (below 100 ms) and long SOAs has to be made (Scheerer, 1973a, b). While integration masking is likely to occur for short SOAs, masks with a longer SOA to the target display have an interrupting influence on visual processing. A trend test on linearity for the capacity estimates for enumeration throughout the forward masking SOAs revealed a significant

effect ($F(1,15) = 82.989$, $p < .001$, $\eta_p^2 = .847$), whereas no such linear trend was observable for the memory task within the same temporal range ($F(1,15) = .088$, n.s., $\eta_p^2 = .006$). In contrast, after the time period in which backward masking has an interrupting influence on the perceptual process (around 100 ms, Scheerer, 1973a, b), only visual working memory capacity increases linearly with longer SOAs ($F(1,15) = 8.222$, $p < .015$, $\eta_p^2 = .354$). Enumeration capacity, however, already reached asymptotic values by 100 ms and showed no further linear effects ($F(1,15) = .877$, n.s., $\eta_p^2 = .055$).

In order to pin down this interaction between task and masking type statistically, we calculated the average performance increase in terms of capacity from the shortest to the longest SOA within the respective time courses of integration (forward masking 24 – 494 ms SOA) and interruption masking (backward masking 118 – 506 ms SOA):

$$(1)\ \Delta K_{forward} = K_{SOA\ 494\ ms} - K_{SOA\ 24\ ms}$$

$$(2)\ \Delta K_{backward} = K_{SOA\ 506\ ms} - K_{SOA\ 118\ ms}$$

These capacity differences were subject to a within-subjects ANOVA with the factors task (enumeration, change detection) and masking type (forward, backward). Both main effects were significant (task: $F(1,15) = 17.207$, $p < .001$, $\eta_p^2 = .534$; masking type: $F(1,15) = 11.888$, $p < .004$, $\eta_p^2 = .442$), More importantly, both factors interacted (task x masking type: $F(1,15) = 66.861$, $p < .001$, $\eta_p^2 = .817$). Whereas enumeration capacity increased over time under forward/integration masking ($M(\Delta K_{forward}) = 1.78$ items, $SD(\Delta K_{forward}) = 0.78$ items; $M(\Delta K_{backward}) = 0.12$ items, $SD(\Delta K_{backward}) = 0.50$ items), vSTM capacity increased over time with backward/interruption masking ($M(\Delta K_{forward}) = 0.02$ items, $SD(\Delta K_{forward}) = 0.71$ items; $M(\Delta K_{backward}) = 0.47$ items, $SD(\Delta K_{backward}) = 0.65$ items).

Summing up, the type of visual masking interacted with the performed task. Interruption masking appeared to exclusively influence the consolidation of information in

visual working memory, with little effect on enumeration. Conversely, increasing the forward mask SOA yielded gradually increasing capacity in the enumeration task, while change detection capacity remained stably poor throughout the whole range of stimulus onset asynchronies.

3.3 Discussion

Overall, the findings are consistent with the hypothesized effect of masking on different stages of object processing (Figure 3.1). These results suggest a close link between capacity limits (in both subitizing and visual working memory) and temporal constraints on object individuation and identification. It adds to extensive empirical and theoretical work that indicates that object file formation involves a temporal succession of processing steps: target detection is faster than target identification in visual search (Sagi & Julesz, 1985), post-offset location information is processed sooner than identity information (Schiller, 1965; Finkel & Smythe, 1973), spatiotemporal information allows an 'object file' to be created, before it is filled in with object features (Kahneman et al., 1992), and spatial locations are pre-attentively indexed first followed by featural information only becoming available later to attention-dependent mechanisms (Pylyshyn, 1994; but see Egeth, Leonard & Palomares, 2008; Olivers & Watson, 2008 for the role of attention in individuation).

This raises the question of why there are different spatio-temporal windows involved in object perception, one reflecting individuation (visual persistence) and one limiting identification (consolidation into vSTM). One possible explanation is that it reflects the brain's strategy to deal with the need to spatially and temporally integrate information coming from a continuous flow of sensory information.

As known from mathematical and engineering sciences, non-linear positive

and/or delayed feedback systems that are engaged in real-time processing exhibit asymptotic unstable behavior when confronted with signals with different latencies that have to be combined (Sandberg, 1963). In such a system, there is a disequilibrium between the need for dynamic and flexible representations (emphasizing new information) and the need for stable and reliable visual representations (maintaining the current state). This trade-off between stimulus read-out and perceptual synthesis can be achieved by temporal multiplexing of feedforward and feedback signals (Öğmen, 1993).

We suggest that this need to balance feedforward and feedback processes must inherently limit capacity for rapid object individuation. According to this model, the real-time dynamics of visual processes unfolds in three phases: (1) afferent feedforward signals allow the read-out of the sensory information; (2) during the decay of the feedforward signal, a feedback or re-entrant dominant phase establishes perceptual synthesis; (3) a reset phase is initiated, resulting in an inhibition of the feedback signals and a re-establishment of the feedforward-dominant mode that delivers the new signal. This succession of transient epochs implements a degree of inertia in the system's response to changes in input and thus limits its real-time dynamics in order to guarantee an equilibrium between flexibility and stability of the visual representation (Öğmen, 1993; Enns & Di Lollo, 2000). Of course, the solution of multiplexing creates temporal windows of visual persistence during which only a limited number of objects can be processed.

In accordance with this idea, we have reported evidence that capacity limits in enumeration depend, at least in part, on a "magic window" of sensory persistence (see also Wutz et al., 2012), which determines the "magic number" of around 4. Using integration masking, the effective persistence of the target display can be fractionated (Di Lollo, 1980; Wutz et al., 2012), thus reducing the effective lifespan of the feedforward dominant phase and thus limiting the time to read-out spatiotemporal object information

and create ‘object files’. This forward masking technique appears to act early in the individuation stage in which targets are segmented and spatiotemporally segregated from the background.

As described above, in a second phase the effective signal strength decays and the system enters the re-entrant phase of processing, during which object identification mechanisms fill in the feedforward established ‘object files’ with featural content. Thus, in addition to the first capacity limit resulting from the effective persistence of the stimulus, a secondary limit comes from the consolidation of information into visual working memory. In particular, this consolidation process can be interrupted if new visual input arrives during the phase of feedback identification processing, since this new stimulus initiates a new feedforward process during this crucial phase of inertia, leading to “interruption masking” (Scheerer, 1973a, b; Enns & Di Lollo, 2000; Breitmeyer & Öğmen, 2006). The speed of object identification, and thus the formation of high-resolution object files, is influenced by processing demands and encoding complexity (Alvarez & Cavanagh, 2004). As demonstrated previously, visual working memory performance rises gradually to asymptote under the influence of long backward masks (Gegenfurtner & Sperling, 1993; Vogel, Woodman & Luck, 2006). Consequently, the over-all capacity of vSTM is limited by temporal buffering both at the feedforward individuation and the re-entrant identification stage of object processing.

The dependence of object individuation capacity on the time window of temporal integration and visual persistence further fosters the central role individuation can play in mediating between the two opposing needs of the visual system in real-time processing: flexibility and stability. The fixed number of newly established ‘object files’ is a direct consequence of the time period of initial feedforward processing, which is a fundamental and computationally inherent characteristic of the temporal dynamics of visual

processing. The information gathered during this constant temporal window enables the organism to preserve basic behavioral potential, like reacting to spatiotemporal changes in the environment by body or eye movements. In order to achieve a more sophisticated interaction with the environment (like identification or memory) and therefore stability on a higher representational resolution, additional processing is necessary at the cost of flexibility to new input.

Although human cognition is remarkably powerful, its online workspace, working memory, appears to be highly limited in the number of informational units it processes (Sperling, 1960; Luck & Vogel, 1997; Cowan, 2000). Here we provide a specific and experimentally testable hypothesis for the origin of cognitive capacity limitations: processing time. Previous proposals about the root of capacity limitations in vision have introduced relatively abstract concepts like “slots” (Luck & Vogel, 1997; Fukuda, Awh & Vogel, 2010) or “resources” (Alvarez & Cavanagh, 2004; Bays & Husain, 2008). While these theories clearly have augmented our understanding of visual object capacity on a descriptive level, our explanation for capacity limits accounts for them in terms of known mechanisms and embeds the ongoing debate about processing limits into the already well-established body of work about the temporal dynamics of the visual system (Wundt, 1899, 1900; Sperling, 1960, 1963; Shallice, 1964; Ullman, 1984; Loftus et al., 1992; Gegenfurtner & Sperling, 1993; Singer, 1999b; Enns & Di Lollo, 2000; Roelfsema, Lamme & Spekreijse, 2000; VanRullen & Koch, 2003; Busch, Dubois & VanRullen, 2009). As also stated above, our explanation is fully compatible with “resource-“ or “slot-based” approaches, but emphasizes a different perspective on the formation of object representations that can be empirically investigated and directly observed in the laboratory.

In practical terms, such an approach would allow for normal or clinically relevant variability in processing capacity to be broken down into concrete factors such as variations in temporal integration periods, speeded-up or slowed-down employment of selective attention or altered read-out slopes of individuation mechanisms. On a theoretical level, we argue that formal descriptions of selective attention and object file formation (Koch & Ullman, 1985; Itti, Koch & Niebur, 1998; Blaser, Sperling & Lu, 1999) should be augmented by a temporal dimension and not solely focus on spatial characteristics of the visual display (Burr, 1984; Burr, Ross & Marrone, 1986; Lisman & Idiart, 1995; Dempere-Marco, Melcher & Deco, 2012). The explanation of object capacity in terms of temporal constraints on the underlying mechanisms fosters the link between space and time, and the role of both of these *a priori* concepts in sensation (Kant, 1899). These two aspects are both fundamental to human cognition, since "space *and* time are the pure forms of (…) sensation."

(Kant, 1899, p. 164; omissions are indicated by (…)).

4. Temporal windows in visual processing:

‘Pre-stimulus brain state’ and ‘post-stimulus phase reset’ segregate visual transients on different temporal scales

The study reported here has been submitted under the above title by Wutz, A., Weisz, N., Braun, C., & Melcher, D. to the Journal of Neuroscience (under review).

Dynamic vision requires both stability of the current perceptual representation and sensitivity to the accumulation of sensory evidence over time. Here we study the electrophysiological signatures of this intricate balance between temporal segregation and integration in vision. Within a forward masking paradigm with short and long stimulus onset asynchronies (SOA), we manipulated the temporal overlap of the visual persistence of two successive transients. Human observers enumerated the items presented in the second target display as a measure of the informational capacity read-out from this partly temporally integrated visual percept. We observed higher β-power immediately prior to mask display onset in incorrect trials, in which enumeration failed due to stronger integration of mask and target visual information. This effect was timescale specific, distinguishing between segregation and integration of visual transients that were distant in time (long SOA). Conversely, for short SOA trials, mask onset evoked a stronger visual response when mask and targets were correctly segregated in time. Examination of the target-related response profile revealed the importance of an evoked α-phase reset for the segregation of those rapid visual transients. Investigating this precise mapping of the temporal relationships of visual signals onto electrophysiological responses highlights how the stream of visual information is carved up into discrete temporal windows that mediate between

segregated and integrated percepts. Fragmenting the stream of visual information provides a means to stabilize perceptual events within one instant in time.

Relevant information from the visual environment can change dynamically due to either real-world transitions (i.e. change or motion) or internal shifts in focus (i.e. spatial attention, eye movements). However, we experience our sensory surrounding to be coherent and stable in time and space (Melcher, 2011). Perceiving visual stability requires an intricate balance between reading-out spatiotemporally invariant representations (i.e. objects) and simultaneously accumulating further sensory evidence over time. Intermediate-level vision has to mediate virtually in real time between segregating individual objects detached from their immediate spatio-temporal reference and integration of sensory flux (Öğmen, 1993; Öğmen & Herzog, 2010; Wutz et al., 2012; Wutz & Melcher, 2013).

We investigated the electrophysiological (MEG) signatures of temporal segregation and integration in vision by presenting observers with two successive sensory signals. Our study took advantage of a forward masking paradigm to manipulate the temporal overlap between two visual transients: mask and target (Di Lollo, 1980; Wutz et al., 2012). The task was enumeration, which unlike simple target detection requires structuring operations in intermediate-level vision (object individuation (Xu & Chun, 2009), visual routines (Ullman, 1984)) whose outputs in form of object files can provide visual stability over time (Pylyshyn, 1989; Kahneman et al., 1992). Critically, the capacity of individuation depends upon the stimulus onset asynchrony (SOA) between target and mask, which determines the degree to which visual persistence of the two stimuli is integrated (Wutz et al., 2012; Wutz & Melcher, 2013). We compared correct trials, in which mask and target were successfully segregated in time, to those in which integration by masking was stronger and enumeration failed.

Electrophysiological signatures of temporal segregation and integration in vision were expected to be predominant in three key time periods (Figure 4.1). First, as suggested by previous paradigms probing the influence of ongoing brain activity (Varela

et al., 1981; Thut, Nietzel, Brandt & Pascual-Leone, 2006; Romei, Brodbeck, Michel, Amedi, Pascual-Leone& Thut, 2008) or top-down control on perception (Hanslmayr, Aslan, Staudigl, Klimesch, Herrmann & Bäuml, 2007a; van Dijk, Schoffelen, Oostenveld & Jensen, 2008; Keil, Müller, Ihssen & Weisz, 2012; Keil, Müller, Hartmann & Weisz, 2013; Volberg, Wutz & Greenlee, 2013), we predicted higher power within the alpha (8-12 Hz) to beta frequency range (13-30 Hz) prior to mask onset for incorrect trials. Second, effects of temporal integration should be observable in the magnitude of the response evoked by the first masking transient (Winkler, Reinikainen & Näätänen, 1993; Hamada, Otsuka, Okamoto & Suzuki, 2001). Finally, we compared target related processing selectively for short and long SOAs. Adding the specific SOA to the latency of the initial mask evoked response provides an estimate of when signals related to visual processing of a target display can be expected on a particular sensor (Rieger, Braun, Gegenfurtner & Bülthoff, 2005). We examined the temporal relationship between these expected and observed responses related to individuating target information from masking persistence in close (short SOA) and distant (long SOA) temporal proximity.

This precise mapping of the temporal relationships of visual signals onto electrophysiological responses allowed us to investigate the role of discrete temporal windows in segregation and integration of visual information, as a means to stabilize vision over time.

4.1 Methods and materials

4.1.1 Subjects

Sixteen participants volunteered after giving written informed consent. Two participants were excluded from analysis: one due to excessive artifacts in the MEG data, which contaminated over 50% of the trials, the other because of exceptionally bad behavioral performance (less than 60 % of correct responses in the easiest experimental condition

(one target item with 200 ms SOA)). Fourteen subjects remained in the sample (eleven female; mean age *M* = 25.1 years, *SD* = 1.9 years; thirteen right handed). All participants had normal or corrected-to-normal vision and took part in exchange for payment. The experimental protocol was approved by the local ethics committee.

4.1.2 Stimuli and procedure

Prior to the experimental runs, each participant completed 50 practice trials to familiarize them with the visual stimulation and the response collection devices. The experimental procedure started only after the mapping between response finger and button box became relatively automatic (at least 20 consecutive fast and correct responses). Visual stimuli were presented to subjects in a dimly lit magnetically shielded room. The visual stimuli were generated on a HP Intel Quad core computer using MATLAB 7.9 (MathWorks, Natick, MA) and Psychophysics Toolbox Version 3 (Brainard, 1997; Pelli, 1997). A DLP projector (Panasonic PT-D7700E, Osaka, Japan) projected the visual stimuli at a refresh rate of 60 Hz centered onto a translucent screen (22° (horizontal) x 17° (vertical) of visual angle), located 127 cm from the subjects. The precise timing of the visual stimulation was monitored via a photo diode placed at the upper left corner of the projection screen and the delay between trigger and stimulation onset was corrected with this method.

Each trial began with a central fixation dot (black, 0.15°) on a white background for 500 ms, followed by a blank white screen for a jittered pre-mask interval (800 – 1300 ms). The visual stimulus consisted of a forward mask and a target display superposed onto the masking pattern. On each trial a different pattern of 2250 randomly oriented, partially crossing black lines (mean line length = 0.5° visual angle, mean line width = 0.04°, mean size of whole pattern = 4° (horizontal) x 5.6° (vertical)) was presented centered on a white background first (Figure 4.1). This pattern remained on the screen

and after a variable onset delay (SOA), from 0 to 4 target items appeared superimposed upon the random line pattern by use of the image processing technique called ‘alpha blending’ (alpha = 0.5 for both displays; Figure 4.1). Two diagonally crossing lines (‘X’) represented one target item. All items were colored in black, were 0.3° (horizontal) x 0.5° (vertical) of visual angle in size and were placed randomly on one of 16 possible locations within an invisible, central rectangle of 2.4° (horizontal) x 3.3° (vertical) of visual angle in eccentricity with a minimum buffer of 0.4° between the locations (Figure 4.1). The physical properties of both mask and target elements, i.e. contrast, mean line length, mean line width, were equated. Furthermore the ‘alpha blending’ procedure edited the transparency/opacity values of the visual stimuli assuring a mathematically correct superimposition of local element contrast, without creating any discontinuities in luminance. All these adjustments assured that mask and target elements only differed in their temporal onset exclusively creating partial overlap between the visual persistence of the two transients (Di Lollo, 1980; Wutz et al., 2012; Wutz & Melcher, 2013).

The target display was presented for 50 ms, the preceding masking pattern, however, was on the screen during target presentation plus the independently varied stimulus onset asynchrony between mask and target display. There were four different stimulus onset asynchronies: 0, 33, 50 or 200 ms (Figure 4.1). After mask and target offset a white blank screen was presented until the subject’s response (which initiated the next trial) or for a maximum of 2 s. The participants’ task was to indicate the quantity of perceived items in the target display by lifting the finger in the corresponding optical fiber button boxes, which were assigned one particular number each before the experiment (5 boxes for responses 0-4). The finger-response mapping was balanced across subjects. In total each participant ran 20 blocks with 102 trials per block (about 6 min duration).

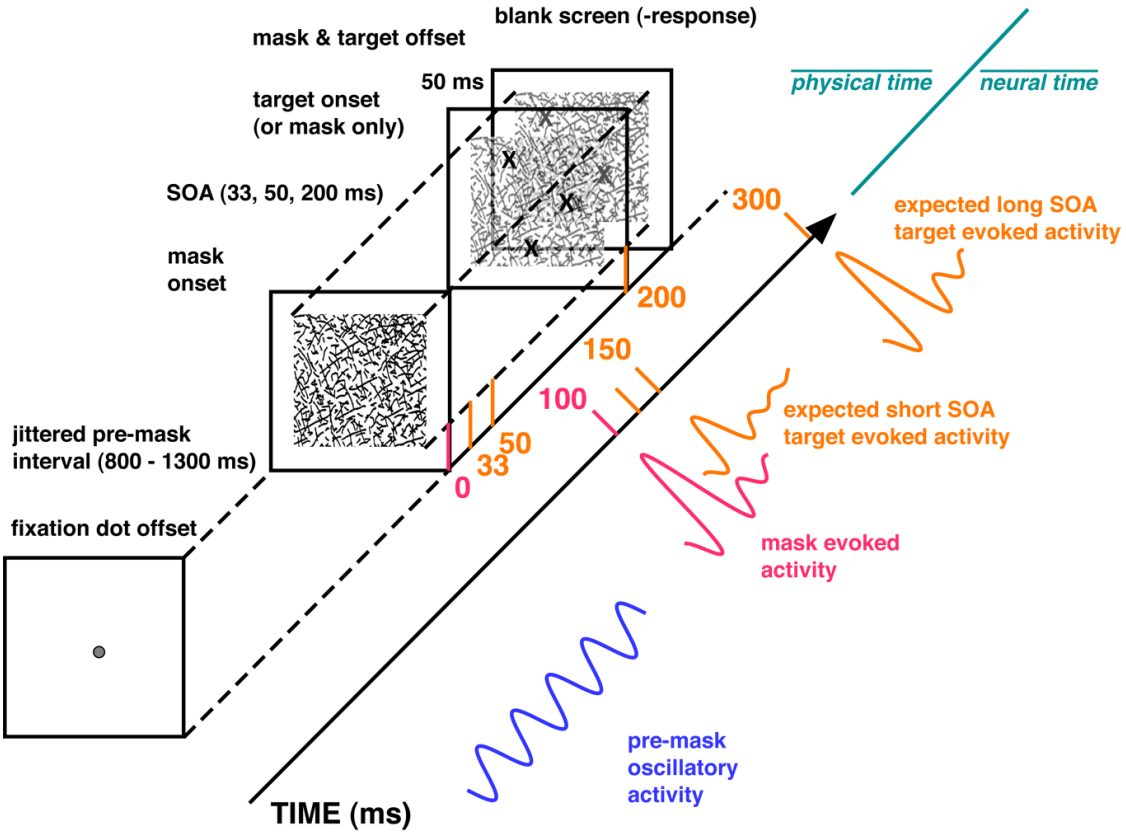


Figure 4.1 Temporal integration masking: stimuli, trial sequence & expected MEG effects

Illustration of a typical sequence of visual stimuli within one trial (here shown for the case of 200 ms of SOA between mask and target display). With the offset of a fixation dot commenced a temporally jittered pre-mask period (800-1300 ms). Then a random line pattern was presented and remained on the screen when the target display (two diagonally crossing lines ('X') with varied set size) was presented. The stimulus onset asynchronies (SOAs) between mask and target onset used in this experiment were 0, 33, 50 & 200 ms. The targets superposed upon the masking pattern (here shown 60% transparent for illustrative reasons) were always shown for 50 ms. Whereas we expect ongoing or cognitively induced oscillatory activity in the pre-mask interval, the onset of the masking pattern is expected to evoke a visual response around 100 ms after mask onset (estimated from the data, although caution has to be taken as to the exact absolute latency of the evoked response (VanRullen, 2011).) Consequently, the expected latency of the target-evoked response given an affine transformation of physical time to neural time can be estimated by adding the specific SOA to these 100 ms. The objections to the absolute latency do not apply to this relational metric.

Each of the 16 possible combinations of SOA (0 - 200) and set size (0 - 3) were presented six times per block in random order. Displays containing four target items were only shown six times and always in the trials with SOA of 200 ms. This set size was included as catch trials to prevent a response bias to always report the highest number when in doubt. Showing target displays of different set sizes is important to measure integration masking per se. However, for the analysis of the electrophysiological activity, we combined the responses for the different set sizes. Since no response bias for the set size of 4 items was evident, we collapsed the data for all sizes and only contrasted correct and incorrect trials irrespective of the actual set size. Moreover, trials with no targets constitute a separate experimental condition. These mask only trials collapsed over the different SOAs serve as a control condition for mask-evoked activity without additional target processing.

4.1.3 MEG measurement

Electrophysiological activity was recorded with an online sampling rate of 1000 Hz using a whole-head MEG with 102 magnetometers and 204 planar gradiometers (Neuromag306 system, Elekta, Stockholm, Sweden) in a magnetically shielded room. This system consists of 102 sensor locations each containing a triplet of one magnetometer and two gradiometers. In particular gradiometer information is sensitive to sources close to the sensor location, i.e. neural generators at the cortical surface. In order to localize the head position of the subject within the MEG helmet, a subject-specific head frame coordinate reference was defined before the experimental runs. The cardinal points of the head (nasion and left and right pre-auricular points), the location of five head position indicator (HPI) coils and a minimum of other 200 head shape samples were digitized for motion tracking (3Space Fastrack, Polhemus, Colchester, Vermont) at the start of each session. The subject's head position relative to the HPI coils and the

MEG sensors was estimated before each experimental run to ensure that no large movements occurred during the data acquisition procedure.

4.1.4 MEG data analysis

The data were analyzed using the FieldTrip toolbox (Oostenveld, Fries, Maris & Schoffelen, 2011) in combination with MATLAB 7.12 (MathWorks, Natick, MA). The data were segmented from 800 ms before to 1500 ms after mask onset, downsampled off-line to 250 Hz and band-pass filtered between 1 and 40 Hz with a two-pass Butterworth filter with the order 4. A semi-automatic artifact detection routine identified trials and channels that deviated in amplitude using a summary statistic (variance) of the entire data set. These trials and channels were removed from the data set. Finally, the data were visually inspected and any remaining trials and channels with artifacts were removed manually. The rejected channels were interpolated with the nearest neighbors approach for sensor level analysis. For source localization information from interpolated channels was not used. Finally the proportions of trials for the experimental conditions of interest (correct, incorrect, mask only trials) were equated in trial number by selecting a random subsample of trials from the condition with more trials.

4.1.4.1 Event-related fields

Before calculating event-related fields (ERFs), data were band-pass filtered using a two-pass Butterworth filter with a filter order of 4 and a frequency cutoff between 2 and 20 Hz. After calculating the event-related field as the average in amplitude across trials, data from planar gradient pairs were combined using vector addition. ERFs were baseline corrected using an interval of -200 to 0 ms before mask onset.

4.1.4.2 Time–frequency analysis

We calculated time–frequency representations (TFRs) using a Fourier transform approach applied to short sliding time windows in steps of 10 ms. The power estimates were computed by means of a Hanning taper of single trial data for the frequency range from 5 to 40 Hz. The window length of the taper was 5 cycles per frequency of interest. This procedure yields good spectral resolution at low frequencies and good temporal resolution at high frequencies. The power values were calculated for the horizontal and vertical component of the planar gradient and then combined via their vector sum.

4.1.4.3 Inter-trial coherence

In order to disentangle the effects of an increase in amplitude and an increase in phase consistency across trials of the visual evoked response, we computed the inter-trial coherence (ITC; Makeig, Debener, Onton, & Delorme, 2004); also called phase-locking factor (PLF); Tallon-Baudry, Bertrand, Delpuech & Pernier, 1996) within the interval of -200 to 500 ms around mask onset and the frequency range between 5 and 35 Hz, with identical Hanning taper characteristics as described for the calculation of oscillatory power. At each time, frequency and sensor sample, the result of the Hanning tapering and Fourier transform for each trial is a complex number with a real and an imaginary part. In order to control for differences in amplitude, the lengths of the complex vectors (representing amplitude and phase) were normalized to one for all trials. Thus, only the information about the phase of the spectral estimate of each trial is taken into account. The extent of phase consistency across trials is quantified by the length of the resultant of these normalized complex vectors along the unit circle. The ITC measure can take values between 0 and 1. A value of 0 represents a random phase angle distribution across trials and a value of 1 indicates perfect synchronization across trials between MEG data and the time-locking events. The ITC values were calculated for the horizontal

and vertical component of the planar gradient and then combined via vector addition.

4.1.4.4 Source localization

A structural MRI image was available for 12 of 14 participants. We co-registered the brain surface from their individual segmented MRIs (Nolte, 2003) with a single-shell head model. For 2 of the 14 participants no individual MRI scan was available. For those subjects we obtained the canonical cortical anatomy from the affine transformation of an MNI-template brain to the subject's digitized head shape. Source activities were projected onto these approximate individual anatomical MRI images and subsequently normalized onto a standard MNI brain (Montreal Neurological Institute (MNI), Montreal, Canada; http://www.bic.mni. mcgill.ca/brainweb) using SPM8 (http://www.fil.ion.ucl.ac.uk/spm) in order to accomplish group statistics and for illustrative purposes. Anatomical structures corresponding to the localized sources of the statistical effects were found using the MNI brain and Talairach atlas (MRC Cognition and Brain Sciences Unit, Cambridge, England; see http://imaging.mrc-cbu.cam.ac.uk/imaging/MniTalairach).

4.1.4.5 DICS beamforming of oscillatory sources

The neural generators of the effects found in the time-frequency domain were identified by means of dynamic imaging of coherent sources (DICS, Gross et al., 2001), a frequency-domain adaptive spatial filtering algorithm. This algorithm has proven to be particularly powerful when localizing oscillatory sources (Liljeström, Kujala, Jensen & Salmelin, 2005). A common spatial filter derived from all trials has been applied separately to the different conditions (correct, incorrect). Based on the sensor level effects, power and cross-spectral densities were calculated for 15 Hz (+/- 3 Hz smoothing) and within -500 to 0 ms relative to mask onset. As the pre-mask activity was

mainly of interest here, source analysis outputs for both conditions (correct vs. incorrect trials) were compared directly without prior normalization.

4.1.4.6 LCMV beamforming of evoked sources

The sources of effects found in the time-series analysis were localized using a linear constrained minimum variance beamformer algorithm (LCMV, (Van Veen, Van Drongelen, Yuchtman & Suzuki, 1997). A common spatial filter based on the signal in all trials has been applied separately to the different conditions (correct, incorrect). The covariance matrix has been derived from the band-pass filtered (cutoff frequencies: 2 to 20 Hz) signal within the time course +50 to +200 ms after mask onset. No baseline adjustment has been applied.

Both the information from the magnetometer and planar gradiometer sensors systems were used for source localization after appropriately adjusting the balancing matrix according to the distance of the gradiometers (17 mm). Separate analysis only using the planar gradiometers yield very similar results (data not shown).

4.1.4.7 Statistical analysis

Oscillatory and evoked visual activity were compared between the conditions by means of nonparametric cluster-based permutation (dependent samples) t-statistics (Maris & Oostenveld, 2007). This procedure effectively controls for the type I error accumulation arising from multiple statistical comparisons at multiple time, frequency and sensor samples. First, clusters of spatio-temporal-spectral adjacent suprathreshold differences (dependent samples t-statistics exceeding $p < .05$, two-sided) were identified. Within one cluster t-values were summed up to reveal a cluster level test statistic. Then, random permutations of the data were drawn by exchanging the data between experimental conditions within the participants. The maximum cluster level statistic was recorded after

each permutation run, revealing a reference distribution of cluster level statistics (approximated with a Monte Carlo procedure of 1000 permutations in the present study). Cluster-level p values were then estimated as the proportion of values in the corresponding reference distribution exceeding the cluster statistic obtained in the actual data. Source level comparisons were calculated using dependent-samples t-tests within the effects of interest identified on the sensor level.

4.2 Results

4.2.1 Behavioral data

The proportions of correct trials within the set sizes and stimulus onset asynchronies of interest were fed into a within-subjects analysis of variance (ANOVA). Consistent with previous results (Wutz et al., 2012; Wutz & Melcher, 2013), enumeration performance improved with increasing SOA within the set sizes (1-3 items) and SOAs (33 – 200 ms) of interest (SOA; $F(2,26)= 89.36$, $p < .001$). In fact, a closer examination of the error distributions revealed that the most frequent incorrect response was to report one item less than was actually presented (≈ 50 % of all incorrect responses). Observers seldom missed detecting the onset of the second target display entirely (erroneous responding 0 targets comprised less than 5% of all trials), but instead failed to converge towards the correct response within the effective persistence of the visual image (supplementary table ST4.1). As noted elsewhere this pattern of results is most likely due to increasing enumeration performance with less temporal integration (and hence better temporal segregation) of mask and target visual information (Scheerer, 1973a; Di Lollo, 1980; Enns & Di Lollo, 2000; Wutz et al., 2012; Wutz & Melcher, 2013). Also the second main effect (set size) - showing better performance with smaller presented numerosities ($F(2,26)= 43.82$, $p < .001$) – and the interaction term were significant (SOA x set size; $F(4,52)= 7.57$, $p < .001$). Replicating previous findings (Wutz et al., 2012), small set size

displays could be efficiently enumerated with short SOAs and already reached asymptotic performance (≈ 90 % correct) thereafter, whereas performance with higher set sizes improved also with longer SOAs, within this ordinal interaction (Table 4.1).

Table 4.1 Behavioral results. Enumeration performance (mean and standard deviation (SD) of proportion of correct trials in %) across subjects (N=14).

SOA (ms)	set size	% correct trials	SD (%)
0	mean (all)	2.8	2.8
33	**mean (all)**	**66.5**	**10.5**
	1 item	77.1	11.6
	2 items	64.9	14.0
	3 items	57.5	11.0
50	**mean (all)**	**77.3**	**9.8**
	1 item	86.6	8.2
	2 items	76.3	11.9
	3 items	69.1	11.2
200	**mean (all)**	**85.0**	**7.5**
	1 item	92.7	5.4
	2 items	89.7	8.0
	3 items	85.4	8.7
	4 items	72.1	14.8
mean (all)	0 items (mask only control)	91.0	6.8

The first row shows below chance performance (20 %) for common onset masking (0 ms SOA) and hence total temporal integration of mask and target information across all set sizes. Then the percentage of correct trials across the different levels of SOA (bold) and for each set size is displayed. The last row depicts performance in the mask-only (0

items) control condition across all SOAs.

Since enumeration performance was not above chance level for the 0 ms-SOA (common onset masking), this condition was not included as a cell in the ANOVA design. The immense difference in the proportion of correct trials between common onset masking (0 ms SOA) and the 33 ms SOA condition (more than 60 %), however, shows what enormous impact a temporal lag as small as 33 ms has on task performance and the entire psychometric function (Table 4.1). Indeed, since detection performance is only marginally impaired at such a small temporal onset asynchrony an enumeration task is required in order to be sensitive to the accumulation of information in this short time frame (Wutz & Melcher, 2013). Moreover, set size 0 trials were not included in the ANOVA design, since there is a fundamental conceptual difference between detecting and individuating physically present target items and detecting the absence of visual targets (Table 4.1).

4.2.2 MEG data

In order to identify electrophysiological signatures of temporal segregation and integration prior and in response to the presentation of visual transients, we first globally contrasted correct and incorrect trials collapsed over the three different SOAs (33, 50 & 200 ms). In a subsequent step we investigated whether those signatures might be time scale specific, yielding different patterns for short (33 & 50 ms) and long (200 ms) SOA trials. Mask only trials served as a control condition associated with processing of a single stimulus.

4.2.2.1 Pre-mask oscillatory power

We started off by comparing differences in oscillatory power between correct and incorrect trials in the frequency range from 5 to 40 Hz averaged over the entire pre- to peri-Mask interval (-500 to +50 ms around mask onset). A cluster of central-to-right occipital parietal gradiometer sensor locations (Figure 4.2 B, C) showed significant negative differences (correct < incorrect) in the lower beta band at around 15 Hz (p < .04; Figure 4.2 A). A closer look on the time course of the differences in oscillatory power revealed two temporal maxima at around -350 and -50 ms prior to mask onset (Figure 4.2 A, E). Due to the inherent temporal smoothing of the Hanning tapering ((1/f)*cpf with f: frequency (15 Hz) and cpf: cycles per frequency (5 cpf)) activity differences as close as -50 ms before mask onset could possibly be confounded by stimulus evoked activity (in the present data around +100 ms after mask onset). However, we confirmed that the reported findings are not dependent on the length of the windowing function, since similar effects were found using a shorter window length (2 cpf, supplementary Figures SF4.1 & SF4.2). DICS beamforming in the pre-mask interval (-500 to mask onset (0 ms)) at 15 Hz (+/- 3 Hz smoothing) suggested that neural generators at the right occipital pole (peak difference t(13)=-4.54; MNI coordinates [34.0 -88.0 0]) and in left ventral occipital to inferior temporal areas (t(13)=-4.2; MNI coordinates [-49.0 -48.0 -9.0]) were involved in this power difference seen at the sensor level (Figure 4.2 D).

Beta power decreased both for correct and incorrect trials with approaching mask onset, indicating that mask onset may have been anticipated by the participants (Figure 4.2 E). This pattern suggests that the observed effect may be due to cognitively induced pre-stimulus activity in opposition to fluctuations in an ongoing occipital beta rhythm. It is noteworthy that we also observed oscillatory power differences in the alpha frequency range (8-12 Hz; see Figure 4.2 A) conforming to previous findings (e.g. Hanslmayr et al., 2007a; van Dijk et al., 2008). The higher power in incorrect trials compared to correct

trials, however, did not reach significance on a cluster level (that controls for multiple comparisons) in the alpha band.

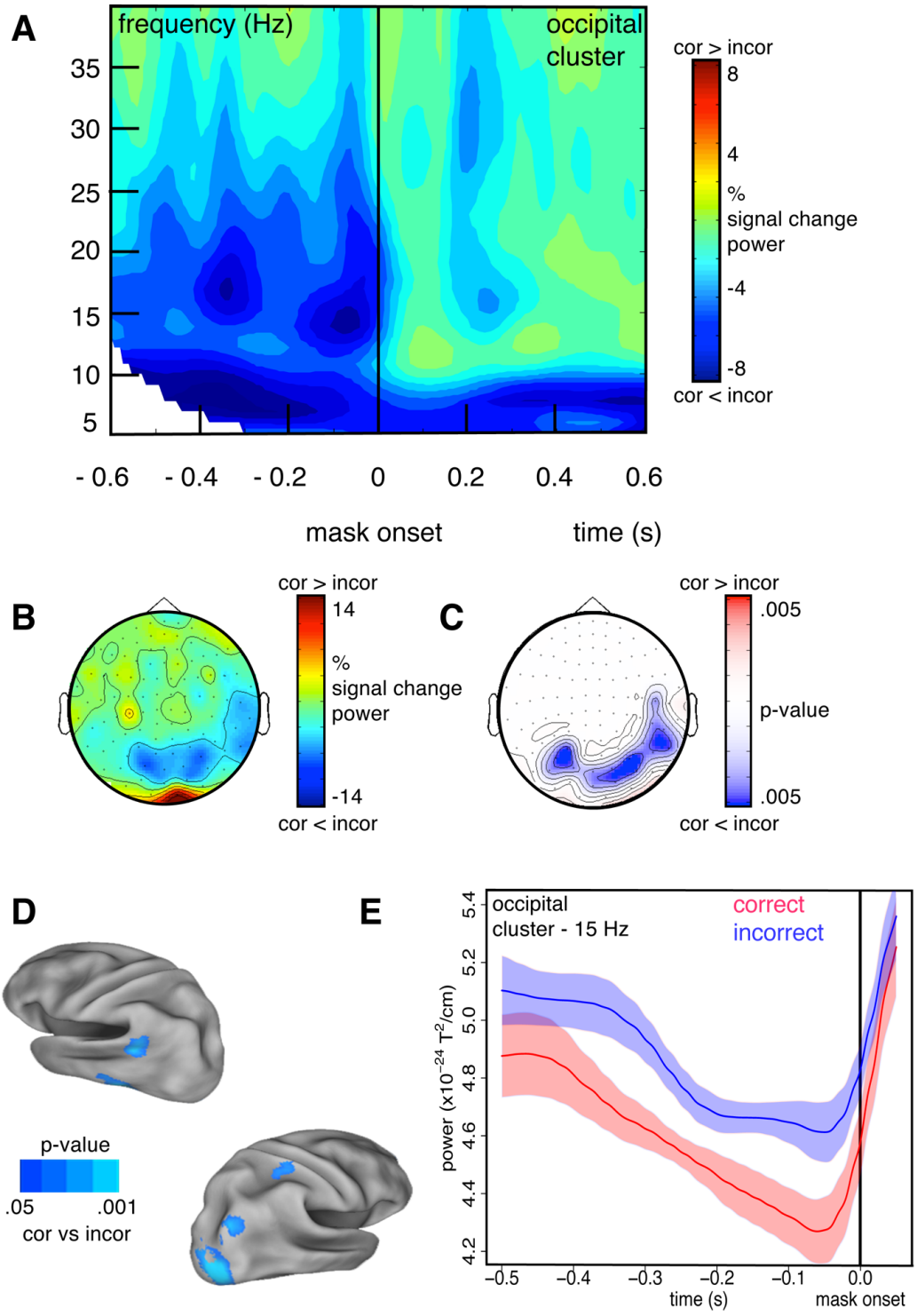


Figure 4.2 Higher pre – mask β power within incorrect compared to correct trials

A Time-frequency plot showing the percent in signal change in oscillatory power within correct in comparison to incorrect trials ((correct-incorrect)/incorrect) around mask onset within an occipital cluster of sensors shown in **B** & **C**. Warm colors indicate higher power in correct trials, cold colors in incorrect trials. There are two obvious effects within lower β frequencies (15 – 20 Hz) at -350 and -50 ms. **B** Head topography showing the percent in signal change in oscillatory power at 15 Hz like in **A** averaged over the 500 ms pre-mask interval. **C** Corresponding head topography on mean power differences at 15 Hz in the 500 ms pre-mask interval between correct and incorrect trials (t-values), non-significant t-values as derived from cluster permutation statistics are masked. Both topographies in **B** & **C** show a cluster of parieto-occipital sensors. **D** DICS-beamformer source localization of the mean differences in oscillatory power (t-values) at 15 Hz averaged over the 500 ms pre-mask interval. T-values below an alpha-level of 0.05 are masked. One oscillatory source is located at the right occipital pole, a second one within left inferior temporal areas. **E** Waveforms showing mean power at 15 Hz within the cluster of sensors depicted in **C** over the 500 ms pre-mask interval for correct (red) and incorrect trials (blue). Shaded areas show the standard error of the mean (solid line) for within-subject designs. Individual performance values have been centered on the mean performance of each subject before calculating the standard error.

Subsequently, we investigated whether the observed effect in oscillatory power in the lower beta frequency band (around 15 Hz) in the interval -500 ms prior to mask onset was time scale specific for the different levels of SOA. Therefore we ran a similar cluster t statistic (frequency of interest from 5 to 40 Hz, gradiometer sensors, averaged over the interval -500 to +50 ms around mask onset) between correct and incorrect trials, now each divided within the three different SOAs (33, 50 & 200 ms). In short SOA trials (33 & 50 ms) no significant clusters of power differences were found (see also Figure 4.3 A & B). For long SOA trials (200 ms), however, a cluster of right occipital sensor locations shows significant differences at 15 Hz ($p < .025$, see also Figure 4.3 C). The general trend of higher beta power within incorrect compared to correct trials in the pre-mask period, however, is observable for all SOA, but strongest for the long SOA trials (200 ms). This effect reaches its maximum immediately before mask onset (-50 ms; Figure 4.3, 4.3 D).

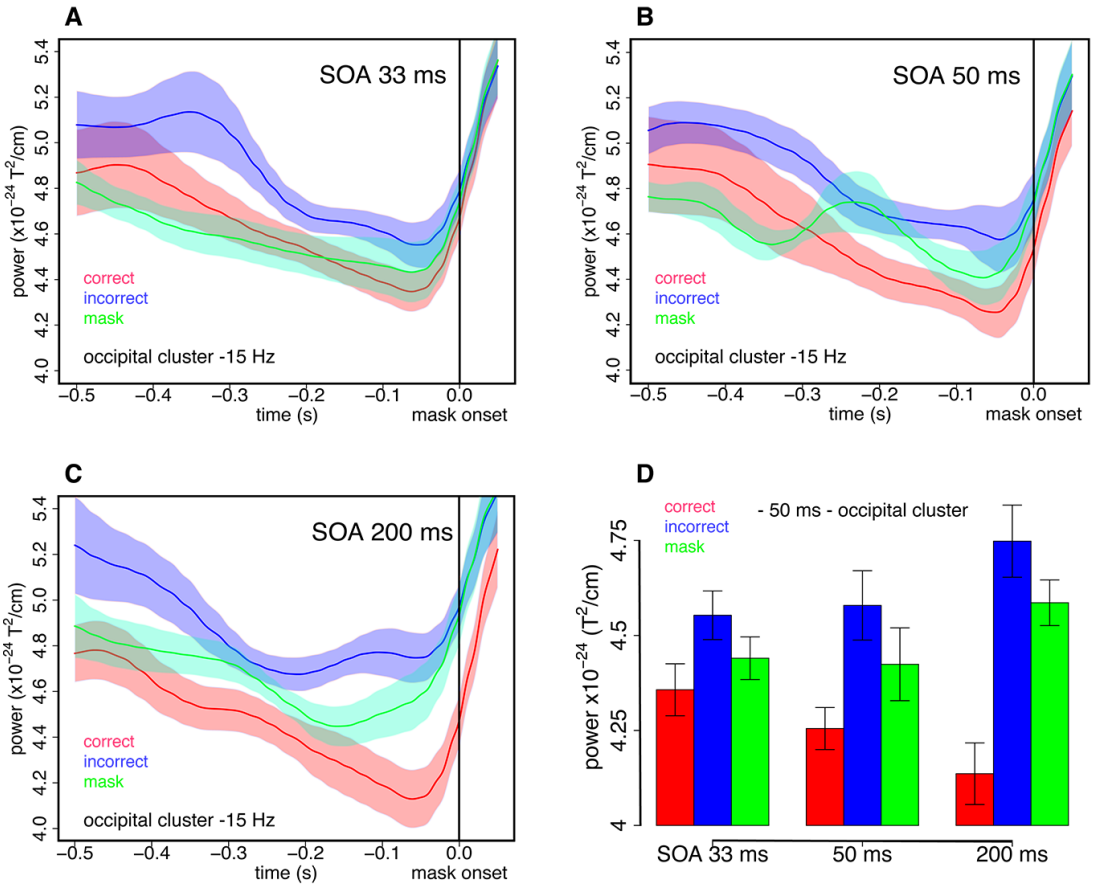


Figure 4.3 Pre-mask β power effect occurs mostly in long SOA trials

A-C Waveforms showing mean power at 15 Hz within the cluster of sensors depicted in **4.2 C** over the 500 ms pre-mask interval for correct (red) and incorrect trials (blue) and mask only trials (green) with a mask-target SOA of 33 (**A**), 55 (**B**) and 200 ms **(C)**. Shaded areas show the standard error of the mean (solid line) for within-subject designs. Individual performance values have been centered on the mean performance of each subject before calculating the standard error. Even though β power is stronger for incorrect trials for all SOAs, this effect is only significant for the long SOA trials (200 ms) on the cluster level and reaches its peak difference around -50 ms before mask onset. **D** Mean power over the occipital cluster of sensors depicted in **4.2 C** at 15 Hz and -50 ms before mask onset for correct (red), incorrect (blue) and mask only trials (green) across SOAs. Error bars show one standard error of the mean for within-subject designs. Individual performance values have been centered on the mean performance of each subject before calculating the standard error. The difference between correct and incorrect trials is strongest for the long SOA trials (200 ms). Both power decreases for correct trials and increases for incorrect trials with longer SOA.

4.2.2.2 Visual evoked response

In the second stage of the analysis, we examined the evoked activity to the two transients, the forward mask and the addition of 1-4 target items. A cluster-based permutation procedure applied on all gradiometer sensor locations and the time interval -500 to +1000 ms relative to mask onset revealed a cluster of central parietal sensors (Figure 4.4 B, C) that showed a significant positive difference ($p < .003$; correct > incorrect) at around 100 ms after mask onset (Figure 4.4 A). The time series of visual evoked amplitude averaged over correct and incorrect trials, respectively, differed significantly around the peak positive deflection (Figure 4.4 A; +60 to +130 ms relative to mask onset). The LCMV source solution yielded relatively widespread activity differences in the interval +50 to +200 ms relative to mask onset onto mostly left hemispheric parietal areas (peak difference $t(13)=6.1$; left inferior parietal; MNI coordinates [-49.0 -55.0 52.0]; Figure 4.4 D).

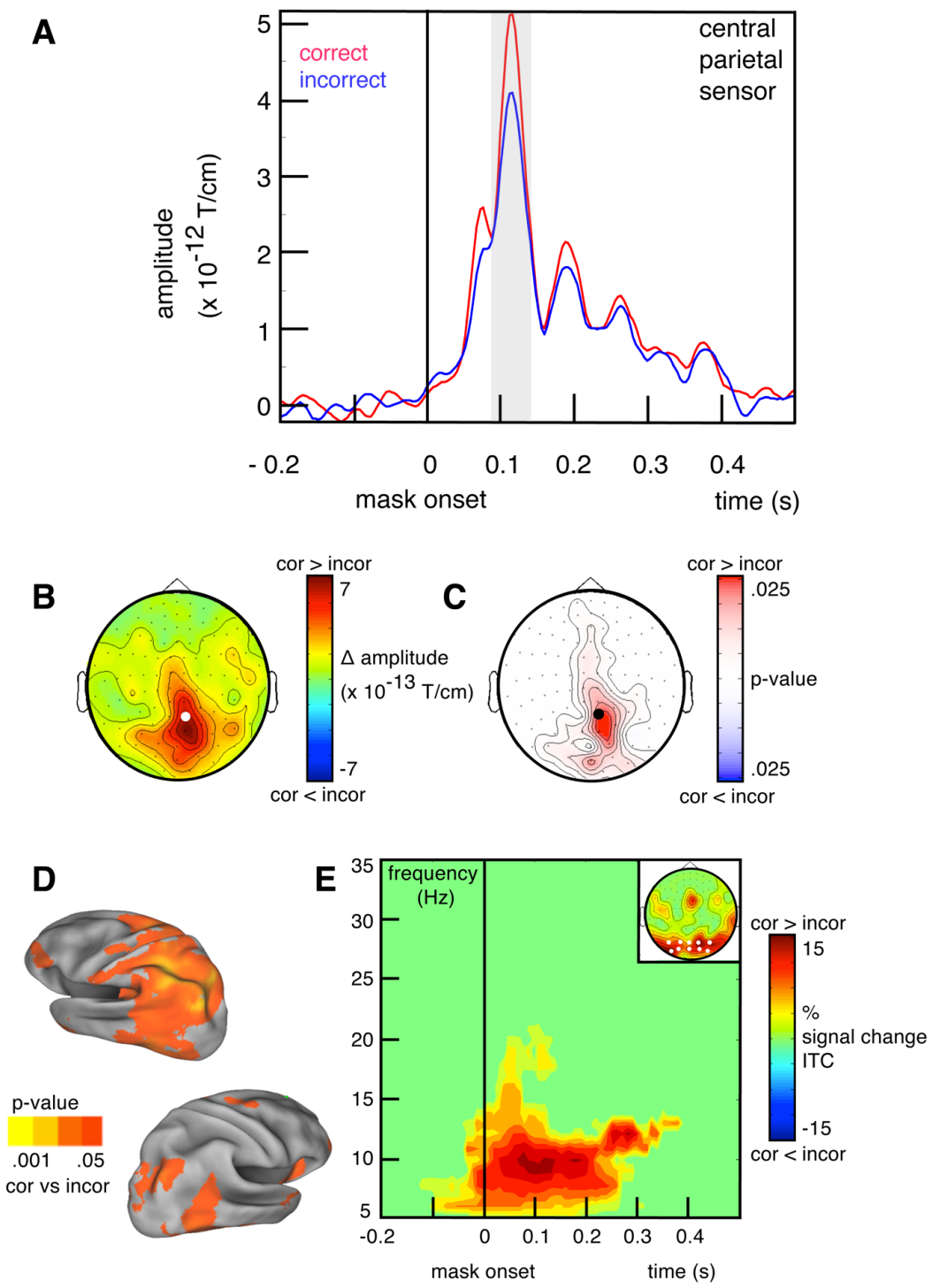


Figure 4.4 Stronger mask evoked response within correct compared to incorrect trials

A Mask evoked response on a representative central parietal sensor (as shown in **B** & **C**). The gray shaded area around the peak response (around +100 ms) denotes the interval within which the visual evoked field differs significantly between correct (red) and incorrect trials (blue) on the cluster level. **B** Head topography showing the difference in

amplitude between correct and incorrect trials averaged over the interval from +60 to +130 ms after mask onset. Warm colors indicate higher amplitude in correct trials, cold colors in incorrect trials. **C** Corresponding head topography on mean amplitude differences averaged over the interval from +60 to +130 ms after mask onset between correct and incorrect trials (t-values), non-significant t-values as derived from cluster permutation statistics are masked. Both topographies in **B** & **C** show a cluster of central parietal sensors and include the representative sensor (white, black dot) from **A**. **D** LCMV-beamformer source localization of the mean differences in amplitude (t-values) averaged over the interval from +50 to +200 ms post-mask interval. T-values below an alpha-level of 0.05 are masked. Mostly left hemispheric, widespread parietal activity differences account for the effect on the source level (peak difference in left inferior parietal areas). **E** Time-frequency plot showing the percent in signal change in inter-trial coherence (ITC) within correct in comparison to incorrect trials ((correct-incorrect)/incorrect) around mask onset averaged over occipital & parietal sites (white dots shown in top inset). Data points with non-significant differences in ITC (as derived from cluster permutation statistics) and significant differences in power (on single sensor level) are masked. Warm colors indicate higher ITC in correct trials, cold colors in incorrect trials. The major effect is centered around +100 ms after mask onset within evoked alpha activity. The head topography of the percent signal change in ITC averaged over the interval from +60 to +130 ms after mask onset and 7 – 12 Hz is shown in the top inset.

4.2.2.3 Inter-trial coherence (ITC) vs. amplitude of the visual evoked field

The higher evoked response to the forward masking event in trials in which the masking effect was weak (correct trials) is counterintuitive, if one assumes a positive linear relationship between response amplitude and masking efficacy. Therefore we tried to estimate to what extent the observed effect is actually due to increases in response amplitude. Theoretically there are two equally plausible explanations as to the generation of measurable differences in the visual evoked field averaged over trials (supplementary Figure SF4.3). First, in one condition the evoking stimulus (or stimuli) could have resulted in higher (or lower) amplitude around 100 ms after stimulus onset in the underlying source of the signal. Second, in one condition the evoking stimulus (or stimuli) could have resulted in a more (or less) consistent phase reset peaking around 100 ms after stimulus onset in the underlying source of the signal. In both cases, the

averaged amplitude over trials would be higher (or lower) in the respective condition.

We estimated the consistency of the phase alignment in response to the external event by computing the inter trial coherence (ITC) for correct and incorrect trials on all gradiometer channels within the frequency range of 5 to 35 Hz and in the interval from -200 to +500 ms relative to mask onset. A cluster of bilateral occipital sensors (Figure 4.4 E; top inset) showed a positive difference ($p < .001$; correct (ITC ≈ 0.9) > incorrect (ITC ≈ 0.7)) in the alpha frequency range (7-12 Hz). The effect spanned the entire period from mask onset up to +300 ms, but peaked at around +100 ms after mask onset (Figure 4.4 E). A similar non-parametric permutation procedure within the same time range and sensors applied on power did not reveal significant differences between correct and incorrect trials on the cluster level (smallest cluster $p > .7$). In fact Figure 4.4 E shows only those spectral-temporal samples significant in ITC on the cluster level ($p < .05$) that did not also show a concomitant significant increase in power on the uncorrected, single sensor level ($p < .05$).

Results indicate that the increased amplitude of visual evoked responses for correct as compared to incorrect trials was at least in parts due to a stronger phase synchronization locked to mask onset in the first condition. Note, however, that we are not making any general claim that the visual evoked field was generated by a phase reset (Makeig et al., 2002; Hanslmayr et al., 2007b), instead emphasize that the observed difference in evoked activity between the experimental conditions of interest in our paradigm (correct and incorrect trials) was to a large extent due to phase consistency over trials and not to an amplitude increase.

4.2.2.4 Visual evoked activity for short and long SOAs

Next, we characterized the temporal dynamics in response to these masking event(s) separately for visual transients in close and distant time intervals (short vs. long SOA). We ran a similar cluster t statistic on the event related field (time series of interest from -500 to 1000 ms relative to mask onset, gradiometer sensors) in correct and incorrect trials, now each divided within the three different SOAs (33, 50 & 200 ms). In long SOA trials (200 ms) no significant cluster of amplitude differences were found (smallest cluster $p > .4$; see also Figure 4.5 C). For short SOA trials (33 & 50 ms), however, a cluster of central parietal sensor locations showed significant differences around 100 ms after mask onset (for 33 ms SOA: $p < .002$ from +70 to +140 ms; for 50 ms SOA: $p < .025$ from +100 to +120 ms relative to mask onset; see also Figure 4.5 A & B). The general trend of stronger evoked amplitude within correct compared to incorrect trials around the peak positive deflection, however, was observable for all SOA, but strongest for the short SOA trials (33 & 50 ms, Figure 4.5 D).

Interestingly the same pattern was observable in purely phase-locked evoked activity measured in inter trial coherence (ITC). Although significant within an occipital sensor cluster in the alpha frequency band (7-12 Hz) for all SOAs (same parameters like in the first analysis step; frequency of interest from 5 to 35 Hz, gradiometer sensors, time series of interest -200 to +500 relative to mask onset; for 33 ms SOA: $p < .001$ from 0 to +300 ms; for 50 ms SOA: $p < .001$ from 0 to +300 ms; for 200 ms SOA: $p < .04$ from +100 to +300 ms relative to mask onset), the effect size between correct and incorrect trials in ITC was almost 7 (for 33 ms SOA) or 4.5 (for 50 ms SOA) times bigger in short SOA trials compared to long SOA trials.

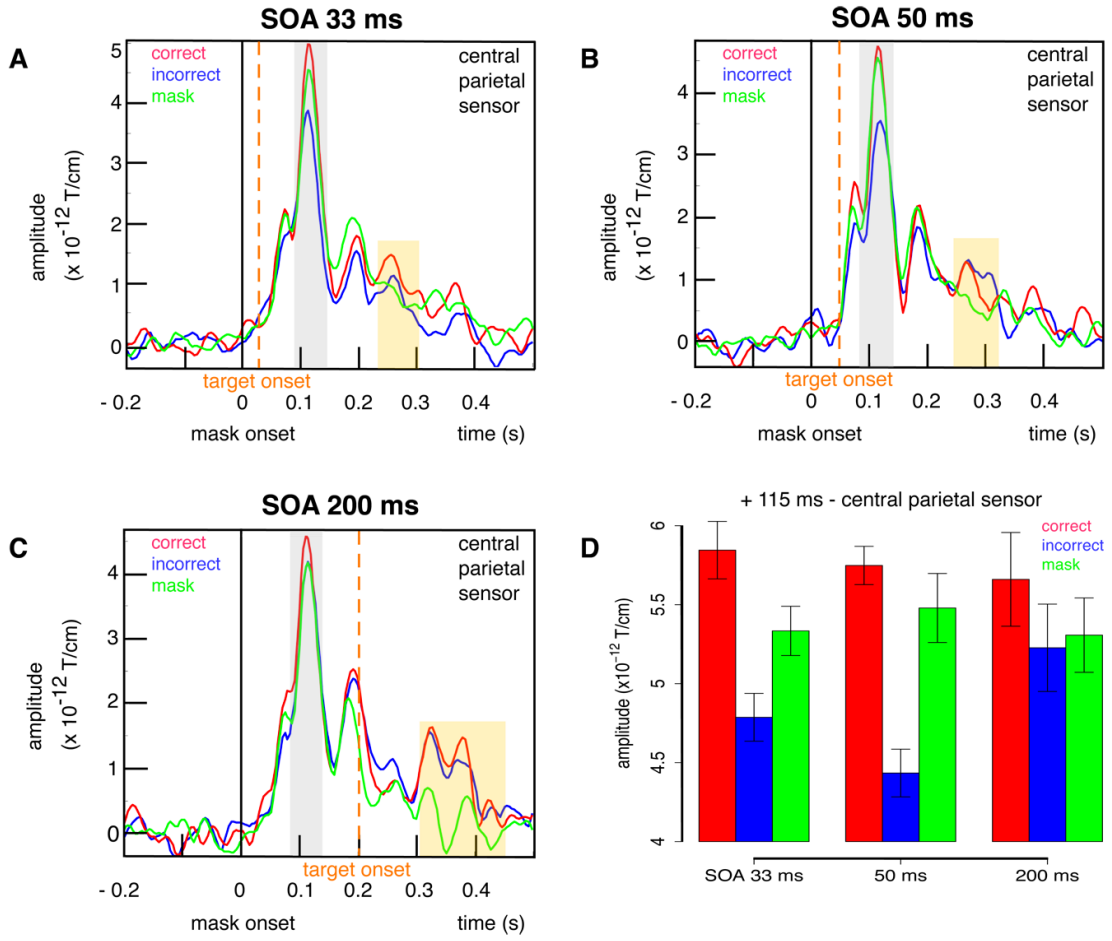


Figure 4.5 Differential evoked response profiles for short and long SOA trials

A Mask evoked response on a representative central parietal sensor (as shown in **4.4 B** & **C**) for trials with a mask-target SOA of 33 (**A**), 55 (**B**) and 200 ms **(C)**. The gray shaded area around the peak response (around +100 ms) denotes the interval within which the visual evoked response differs significantly between correct (blue) and incorrect trials (red) on the cluster level. Even though the evoked response is stronger in correct trials for all SOAs, this effect is only significant for the short SOA trials (33 & 50 ms) on the cluster level The orange shaded area highlights a target related response over an interval during which both correct and incorrect trials differ significantly on the cluster level from mask only activity. The exact latency of this effect varies with SOA: for 33 ms SOA from +230 to +300 ms, but not very visible on this particular sensor; for 50 ms SOA from +260 to +330; for 200 ms SOA from +320 to +460 ms. For better comparability of SOA related latency differences also target display onset is indicated with an orange dotted line for each SOA. **D** Mean amplitude at the representative central parietal sensor (shown in **4.4 B** & **C**) at +115 ms after mask onset for correct (red), incorrect (blue) and mask only trials (green) across SOAs. Error bars show one standard error of the mean for within-subject designs. Individual performance values have been centered on the mean performance of each subject before calculating the standard error. The difference between correct and incorrect trials is strongest for the short SOA trials

(33 & 50 ms). Whereas amplitude in correct trials is relatively stable across SOAs, it is the visual evoked response in incorrect trials – in particular with short SOAs (33 & 50 ms) - that is reduced.

4.2.2.5 Target-related response profile

Evidence for time scale selectivity in the evoked response is further fostered by a differential target related response profile between short and long SOA trials. Evoked magnetic fields of mask plus target trials (correct and incorrect trials) deviated from control mask-only trials consistently across SOAs in a later time window that varied with SOA (Figure 4.5 A-C). Due to this temporal dependency of this effect to target onset, it was probably not observable in the more global first analysis step. For 33 ms SOA, correct and incorrect trials differed significantly from mask-only activity in a left occipital cluster of sensors within the interval from +230 to +300 ms relative to mask onset (p < .001; Figure 4.5 A, but not very visible on this particular sensor). For 50 ms SOA, an occipital cluster of sensors showed this effect within the interval from +260 to +330 ms relative to mask onset (p < .02; Figure 4.5 B). For 200 ms SOA, this effect was located in an occipital parietal cluster of sensors within the interval from +320 to +460 ms relative to mask onset (p < .001; Figure 4.5 C).

The latency of the mask-evoked response (+ 100 ms) can serve as a temporal reference of when evoked responses in general are supposed to arrive at this particular cluster of sensors (Figure 4.1). Thus adding the specific SOA to this reference provides an estimate of when signals related to target processing can be expected (Rieger et al., 2005). In long SOA trials (200 ms), expected (100 +200 ms) and observed (320 ms after mask onset) latencies match quite well. It is important to note that no other significant indicator of only target-related activity was found until 1 s after mask onset on the cluster level within long SOA trials (smallest p > .1). In contrast, for short SOA trials (33 & 50 ms) there is a temporal delay in target related responses between expected (100 +33/50

ms) and observed values (230 /260 ms after mask onset). A response to short SOA target displays is first measurable +100 ms later than expected, if physical time would map affine onto electrophysiological time (Figures 4.1 & 4.5 A-C).

Target displays with a longer SOA also evoked a stronger visual response quantified in cluster effect size compared to short SOA trials (almost 4 times bigger than in 33 ms and almost 9 times bigger than in 50 ms SOA trials). This tendency of decreased attenuation of the evoked response to the second stimulus with increasing onset asynchrony between two successive stimuli is commonly observable within paired stimulus paradigms (Hamada, Otsuka, Okamoto & Suzuki, 2002). The findings are consistent with the characterization of the evoked response as an indicator of the lifetime of sensory memory, based on the attenuation profile of somatosensory responses (Wikström et al., 1996; Hamada, Kado & Suzuki, 2001; Wühle, Mertiens, Rüter, Ostwald & Braun, 2010).

4.2.2.6 Interaction between pre-mask oscillatory power and visual evoked response across SOA

Signatures of temporal segregation and integration can be found both in pre- and post-stimulus intervals. A more thorough examination of the data, however, reveals that these signatures are time-scale specific. Whereas differences in oscillatory power prior to mask onset occur mainly within long SOA trials (200 ms), effects in evoked activity can foremost be found within short SOA trials (33 & 50 ms). In order to pin-down this interaction statistically, we directly compared the effect sizes between pre- and post-stimulus effects across short (33 & 50 ms) and long SOAs (200 ms). Because of the large differences in magnitude between pre-stimulus and evoked activity ($x10^{12}$), the data points of interest (power within occipital cluster of sensors (shown in Figure 4.2 C) at 15 Hz and -50 ms before mask onset and amplitude at central parietal sensor (shown

in Figure 4.4 B & C) at +115 ms after mask onset; Figures 4.3 D & 4.5 D), were standardized to bring them on a common scale (z-scoring). First, we calculated for each subject its average pre- and post-stimulus activity across correct/incorrect trials and different SOAs at the data points of interest. Based on these individual values we estimated the mean (M) and its standard deviation (SD) across subjects separately for pre- and post-stimulus activity and re-referenced each data point to its corresponding sample estimates (z-score: $z = (x_{i,j} - M_{:,j}) / SD_{:,j}$; e.g. $z_{subj1,pre} = (x_{subj1,pre} - M_{:,pre})/SD_{:,pre}$). Then effect size was calculated separately for pre- and post-stimulus effects as the difference of these z-scores between correct and incorrect trials for each SOA. Since pre- and post-stimulus effects have different algebraic signs ($z_{cor,pre} < z_{incor,pre}$; $z_{cor,post} > z_{incor,post}$), the reference category had to be inverted for pre- and post-stimulus effect size calculation in order to yield effect sizes of equal direction (Δ z-score$_{pre} = z_{incor} - z_{cor}$; Δ z-score$_{post} = z_{cor} - z_{incor}$). A 2 x 3 within-subjects ANOVA (pre-/post-stimulus effect x 3 levels of SOA) revealed no significant main effects (pre/post-stimulus: $F(1,13) = .68$, $p > .4$; SOA $(2,26) = .45$, $p > .6$). More importantly, however, the interaction term yielded a highly significant effect (pre/post-stimulus x SOA $F(2,26) = 4.71$, $p < .018$). As expected, strong effects for pre-mask oscillatory power were mainly found within long SOA trials (200 ms). Conversely, short SOA trials (33 & 50 ms) showed pronounced effects particularly for evoked activity (Figure 4.6).

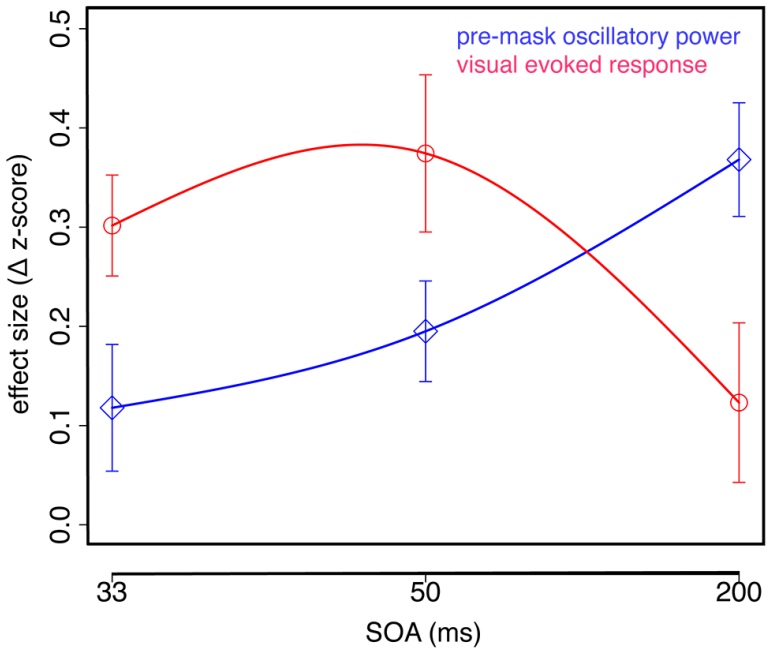


Figure 4.6 Interaction between pre-mask oscillatory power and visual evoked response across SOA

Difference in effect size between correct and incorrect trials (Δ z-score) for pre-mask oscillatory power and visual evoked activity for the three levels of SOA (33, 50 & 200 ms). The z-standardized difference in oscillatory power (blue) between correct and incorrect trials within the occipital cluster of sensors depicted in **4.2 C** at 15 Hz and -50 ms before mask onset (as shown in **4.3 D**) is largest for long SOA trials (200 ms). In contrast, the z-standardized difference in amplitude (red) between correct and incorrect trials at the representative central parietal sensor (shown in **4.4 B** & **C**) at +115 ms after mask onset (as shown in **4.5 D**) is largest for short SOA trials (33 & 50 ms). Error bars show one standard error of the mean for within-subject designs. Individual performance values have been centered on the mean performance of each subject before calculating the standard error.

4.3 Discussion

We found two main signatures of temporal segregation and integration mechanisms in our paradigm – pre-mask β-oscillatory power and the evoked α-phase-locked component of the visual response to transient onset. These seem to be two relatively independent pre- and post-stimulus effects that can be distinguished based on their comparative contribution in segregating temporally close (< 100 ms) or farther apart (200

ms) visual transients. Considering also the third critical time period of evoked responses to the target, short and long SOA trials can be further distinguished based on latency differences. Visual responses to targets in close temporal proximity to a previous mask (short SOA trials) were delayed by approximately 100 ms compared to responses to targets presented later in time (long SOA trials). Within this 100 ms window the consistency of the phase within α oscillations (so approximately within one cycle) was indicative of correct or incorrect performance, but only for short SOA trials. On long SOA trials, on the other hand, modulations in pre-mask β-oscillatory power were associated with task performance.

4.3.1 Pre-stimulus oscillatory power

Incorrect trials, in which mask and target information were more closely integrated into a single percept, showed strong lower β-oscillations (around 15 Hz) throughout the entire pre-stimulus interval (Figure 4.2 & 4.3). Periods of strong β-power have been implicated in top-down control in integration of multisensory signals (Keil et al., 2012, 2013) and spatial contour elements (Volberg et al., 2013). A key attribute of integration across different domains is that local units are combined into a common perceptual or cognitive set (Wertheimer, 1923; Field, Hayes & Hess, 1993). β-oscillatory activity, in particular, has been linked to computational states of maintenance or persistence of the current perceptual/cognitive set, in opposition to a bias towards enhanced sensitivity to new information and expectancy of change (Engel & Fries, 2010).

In temporal vision integrating successive processing iterations within the current perceptual set can help to provide coherence and continuity of the sensory environment. Both correct and - but less pronounced – incorrect trials reveal a clear tendency of decreased power at 15 Hz frequency with approaching mask display onset, and therefore increasing anticipation of perceptual change (Figure 4.2 & 4.3). Integration of

sensory information within the current perceptual set via strong β-oscillations might be the default state of the visual system to emphasize continuity.

Within correct segregation trials, β-oscillations were effectively down regulated time-locked immediately before mask display onset (Figure 4.2 & 4.3). This temporal profile suggests that participants were able to predict the impending onset of the stimuli, enabling the observers to induce a specific neural state to exert top-down control. Classically, β-oscillatory regulation has been reported within sensory-motor tasks like motor imagery (Bai et al., 2008; Waldert et al., 2008), voluntary movement control (Pogosyan, Gaynor, Eusebio, & Brown, 2009) and anticipatory perceptual decision-making (Donner, Siegel, Fries & Engel, 2009). But regulating β-oscillations has in particular been implicated in top-down control in perceptual and cognitive operations (for review see Engel & Fries, 2010), including predictive coding of upcoming perceptual events (Roelfsema, Engel, König & Singer, 1997; Bastos et al., 2012), visual search (Buschman & Miller, 2007), perceptual change in bi-stable images (Okazaki, Kaneko, Yumoto & Arima, 2008) and ambiguous auditory sounds (Iversen, Repp & Patel, 2009). In the current paradigm top-down regulation of β-oscillations could signal the observers' anticipation of the upcoming sensory change. This induced sensitivity to new information, however, has a limited temporal resolution. Induced oscillatory amplitude regulations prior to stimulus onset determine whether temporally distant visual transients (long SOA trials) are segregated or integrated, but play a negligible role for short SOA trials (< 100 ms). If the sensory signal has strong bottom-up constraints, like in short SOA trials in which (temporal) proximity serves as a strong integration cue (Feldman, 2001; Elder & Goldberg, 2002), predictive top-down regulation is deemed ineffective. For such fast visual transients more fine-grained temporal coding is needed.

4.3.2 Post-stimulus phase reset

As a second main signature of temporal segregation and integration, we observed effects in the visual evoked field at around 100 ms after mask display onset. This effect was characterized by stronger phase consistency within approximately one α-cycle (from 100 to 200 ms after mask onset) for correct compared to incorrect trials (Figure 4.4 E). Its observed bandwidth matches psychophysical measures of the effective duration of integration masking (Scheerer, 1973a; Enns & Di Lollo, 2000; Breitmeyer & Öğmen, 2006; Wutz et al., 2012), which in turn reflects the trace a visual stimulus leaves in iconic memory (Sperling, 1963; Di Lollo, 1980; Loftus et al., 1992). In fact strong phase locking is particularly important for segregation of fast visual transients, whose traces overlap within this temporal window, as in short SOA trials (Figure 4.1 & 4.5). Exact phase information within this integration window may be key to allow correct individuation of target information from temporally overlapping masking persistence.

Perturbations in phase consistency could either be a consequence of mask onset alone – acting as a reset event – or interactions between both mask and target transients. A strong reset event in close temporal proximity might ignite an informational trace with higher temporal resolution in which evidence can be accumulated more efficiently (Dehaene, 2011; Zylberberg, Dehaene, Roelfsema & Sigman, 2011). Recently, evidence for the importance of a reset event on subsequent psychophysical α-oscillations has been established behaviorally (Landau & Fries, 2012). Likewise, the strong forward mask might have biased the phase to the active state of the damped oscillation when target evoked signals are supposed to arrive (as shown for ongoing oscillations prior to stimulus onset in target detection (Busch et al., 2009; Mathewson et al., 2009).

Alternatively, the difference in phase consistency across trials might reflect interaction between mask and target. Within a similar paradigm, but together with

backward masking, visual evoked signals sum highly non-linear on the scalp topography, especially when mask and target follow in close temporal succession (Rieger et al., 2005). Hence, interactions of signals from transients in close temporal proximity might perturb phase-locked responses. Phase information has been theorized to be important in connecting coupled systems – like visual signals and visual systems - through coherence (Fries, 2005) or synchrony (von der Malsburg & Schneider, 1986; Singer, 1999a; Engel, Fries & Singer, 2001). In particular visual saliency has been associated with a translation into a phase code via timed release of inhibition (VanRullen & Thorpe, 2001; Klimesch, Sauseng & Hanslmayr, 2007; Jensen, Bonnefond & VanRullen, 2012). Exact phase coding around transient onset may therefore provide a precise temporal integration window within which structuring and individuation of the sensory image relies on this inhibitory timing to accurately encode visual information. Perturbations to this mechanism mediate between segregated and integrated mask-target percepts.

4.3.3 Temporal integration windows

Our perceptual impression reflects the need to construct stable and coherent objects and scenes while also remaining sensitive to new information with high temporal resolution (Melcher, 2011). Given that sensory input arrives continuously, the visual system must mediate between stable and flexible representations virtually in real time (Öğmen & Herzog, 2010). Here we show that when the sensory environment changes rapidly, as in short SOA trials, segregation of these changes depends on precise phase coding within a brief temporal window. Conversely, temporal segregation of sensory changes exceeding this critical time frame depends on slower power modulations prior to stimulus onset.

In contrast to previous studies on the temporal dynamics of target detection, we took advantage of a more sensitive enumeration task in order to probe the early structuring computations within the sensory image (object individuation; (Xu & Chun, 2009); visual routines; (Ullman, 1984)), whose outputs can provide visual stability over time by indexing salient items (Pylyshyn, 1989; Kahneman et al., 1992). When two visual transients are presented in rapid succession, their persistence is partly integrated and thus the time to access the sensory trace of each single stimulus is reduced. In this way, temporal integration limits the computational capacity of individuation of multiple items from a single iconic trace (Wutz et al., 2012). The current study provides evidence that these mask-target interactions occur within a rapid temporal integration window (≈ 100 ms) that maintains the trace of visual persistence.

Precise phase coding within this integration window (through e.g. eigenfrequency damped oscillations (Buzsáki & Draguhn, 2004)) could provide the high temporal resolution necessary for stability of the perceptual representation despite rapid sensory changes. The transmission of a continuous signal into discrete individual entities, however, is necessarily limited by the bandwidth of the carrier function (Shannon, 1948). Thus, such temporal windows constrain the real-time dynamics of visual processing, but likewise offer an explanation for its limited informational capacity (Sperling, 1963; Cowan, 2000). Fragmenting the continuous stream of visual information into different windows of temporal integration provides a neuronal mechanism to maintain the equilibrium between the competing challenges of providing fine temporal and informational resolution of the environment, stabilizing vision within a perceptual instant of time.

5. General discussion about the relationship between temporal and informational resolution in vision

Cognitive functions operate upon a limited number of invariant objects that provide stable information about the external surrounding. In order to arrive at these invariant representations, however, sensory evidence has to be accumulated over a temporal interval until activation reaches threshold (Dehaene, 2011). For online vision in real-time, this interval is particularly crucial since object motion or eye movements induce dynamic changes in the retinal input. Stability and continuity of visual impressions in space and time depends upon an intricate balance between accumulation of sensory evidence over time and ignition of perceptual structure at one particular moment (Melcher, 2011). The reported studies provide evidence for a temporal windowing mechanism in vision that establishes equilibrium between temporal and informational resolution.

In this final section, I will explain how one of the most fundamental enigmas in visual cognition, its limited informational capacity (Cowan, 2000), inevitably follows from incremental accrual of sensory information within a temporal window of finite duration. I will argue that such a temporal window reflects the opposing needs of real-time vision between segregation of incoming information into separate representations and integration of successive input to accumulate sensory evidence over time. Conclusively, I will propose a functional implementation of this temporal window whose efficacy relies on synchrony between establishing a spatio-temporal reference of and encoding individual information within the sensory image, as one perceptual moment of time.

5.1 The temporal window bandwidth determines visual object capacity limits

Humans perceive the external world as organized and structured percepts of objects and scenes (Treisman & Gelade, 1980). In particular, incremental object representations, like object files (Kahneman et al., 1992), provide an interface between external reality and internal concepts. In this way cognition is grounded in the physical world and freed from an infinite regress of referring to semantic categories (Pylyshyn, 2001, 2007). The full set of individual, identifiable and durable objects provides a measure of capacity of the visual online workspace, working memory (Luck & Vogel, 1997). Although human cognition is remarkably powerful, the information within visual working memory appears to be highly limited with a maximum "magical number" of around four informational units (Cowan, 2000).

I investigated the temporal dynamics of individuation as a process in which a complex scene containing individual object-files races to emerge from a uniformly sampled image. Such intermediate-level object files are important in visual cognition since more elaborate identification and consolidation acts upon the individuated data. Fractionating the access to sensory input, I show that unlike transient sampling, individuation capacity increases in steps to arrive within the inherent limits of the current perceptual representation (Figure 5.1).

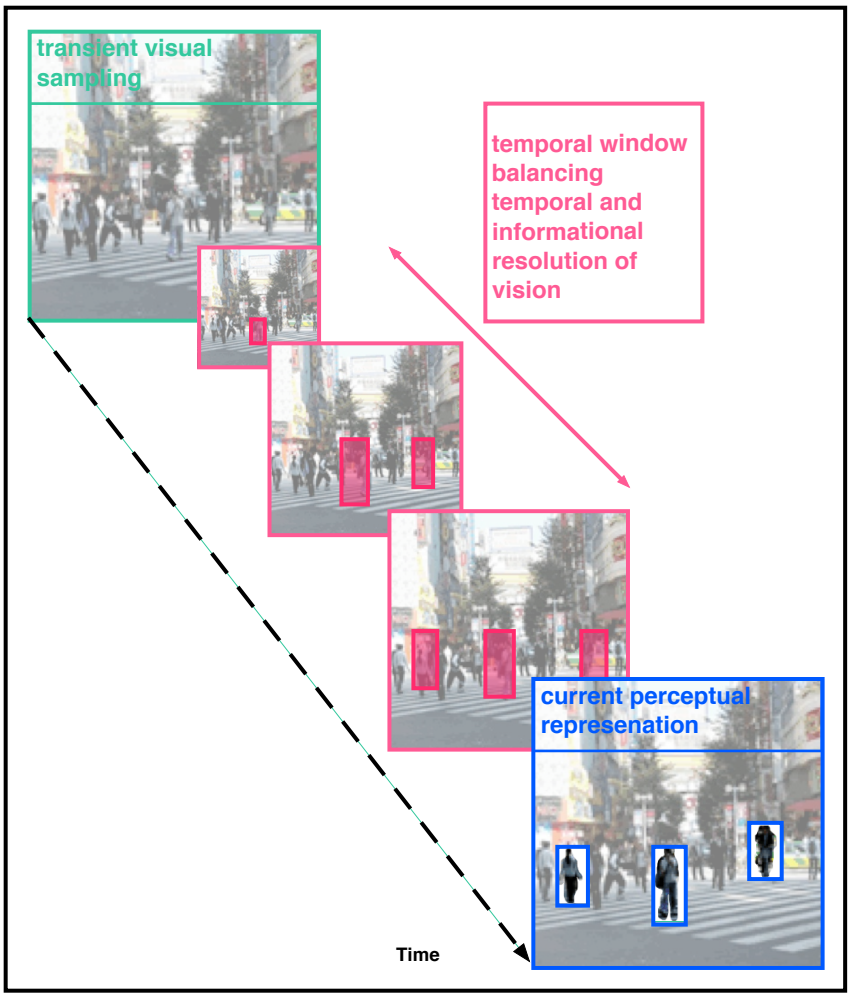


Figure 5.1 Temporal evolution of visual object processing

First, the sensory image is registered. Information on this processing level is high in capacity but easily overwritten (Sperling, 1960). Within a temporal window the sensory image is accessible to intermediate-level structuring operations that successively stabilize the current percept. Stability increases in steps until it converges within its capacity limits.

Consequently, previous theories assuming that vision extracts four informational units instantaneously might have under-sampled the temporal dynamics of individuation. Instead, individuation capacity, as the initial set-up process for object representations, grounds within the bandwidth of a temporal window of sensory persistence that balances between temporal and informational resolution of intermediate-level vision.

This capacity limitation arises from essential sequentiality in processing, in opposition to non-essential sequentiality to merely prevent information overflow (Kahneman, 1973; Holtzman & Gazzaniga, 1982). Sequentiality in information processing is essential when the relevant task requires iterative solutions (Minsky & Papert, 1969). Subjective phenomena such as simultaneity and successiveness entail both stability and continuity of visual impressions virtually in real time. In order to accomplish both tasks simultaneously, the stream of sensory information is hierarchically subdivided into different temporal windows (for review see Pöppel, 1997). Specialized operations (like individuation, identification and consolidation) act upon the sensory information within each window and its outputs are transferred to the next processing step during online visual analysis. Here, I focus on visual computations within the temporal window of sensory persistence. Visual information within this temporal window is conceptualized to represent a meta-state between temporal and informational resolution that allows for the perception of spatio-temporal invariance and change of location over time within the sensory signal.

Individuation can provide stability of the current percept by indexing salient items (Pylyshyn, 1989). The number of individuated object-files therefore measures how much the sensory signal is perceived as an invariant percept at one particular location in time. The higher the capacity of the current perceptual representation, the more are visual images seen as a series of separate instances (Figure 5.2).

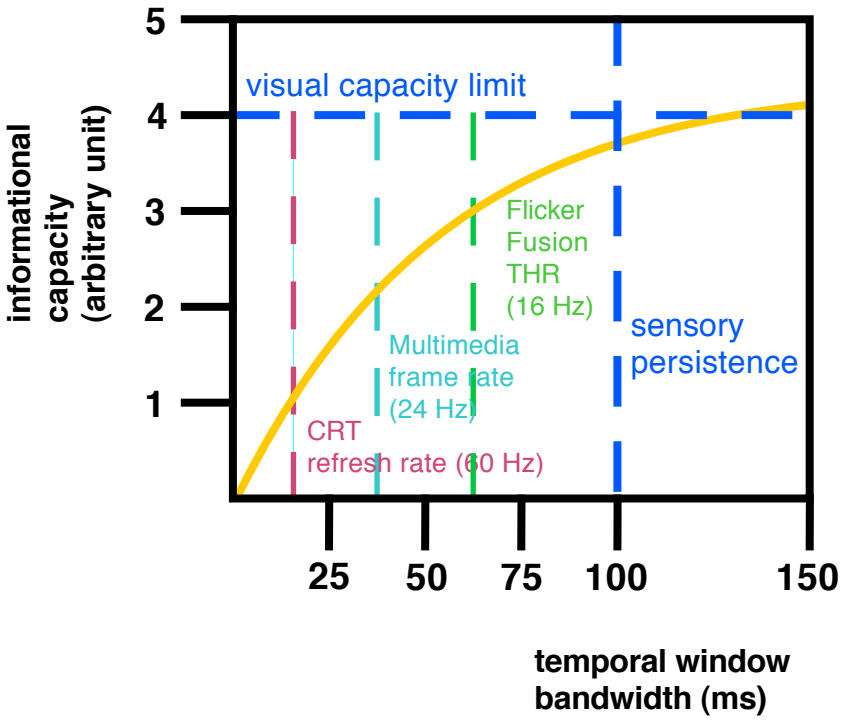


Figure 5.2 Relationship between the bandwidth of sensory persistence and individuation capacity

Visual information measured in individuated object-files increases with decreasing temporal resolution. Visual capacity limits are reached at a bandwidth that matches visual persistence (Di Lollo, 1980). Cathode ray tube (CRT) refresh rates (60 Hz) are barely noticeable, but already allow for single target detection. With a multimedia frame rate (24 Hz) two units can be segregated. Such a temporal resolution is just sufficient to induce beta movement (necessary for apparent motion perception, Wertheimer, 1912). Flicker fusion thresholds vary extensively between individuals, but from around 16 Hz flashed images tend to be seen as separate instances (Crozier & Wolf, 1941).

Within a static surrounding visual capacity limitations unfold for the time of sensory persistence of the stationary percept. If the visual system, however, were coding dynamic events instantaneously into spatio-temporally invariant representations, humans would perceive the external surrounding as a rapid succession of stable, but transient and unrelated images. Such a state might be connected to early neurologic observations about hallucinatory time and speed perception in which the perception of

successiveness is preserved but at higher speeds than physical reality ("Zeitrafferphänomen", time grabbing phenomenon, Hoff & Pötzl, 1934). In contrast, the impression of successiveness would be entirely lost if visual signals would not be translated at all into non-retinotopic and hence spatio-temporally invariant perceptual form (Ögmen & Herzog, 2010). Fast changes would result in motion blur due to incoherently attached visual persistence in space and time (Ögmen, 2007).

Instead, within a dynamic visual environment, the speed of stable information accrual is critical for spatio-temporally coherent vision. Temporal buffering of visual signals provides a means to adjust information accumulation to the needs of real-time vision. In this way, visual computations within one perceptual instant can be conceptualized as an equilibrium between two disparate perceptual situations (Figure 5.3):

Figure 5.3 Opposing needs of temporal and informational resolution in dynamic vision
Dynamic vision has to mediate between two opposing perceptual situations. In the left panel, the object in motion (blurry scooter in the foreground) changes its location faster than camera sampling. Hence successive iterations along its motion path are collapsed to preserve spatio-temporal reference (clear background). In the right panel, the observer focuses on the object in motion (clear scooter in the foreground) to accumulate sensory information, but detaches the object simultaneously from its immediate spatio-temporal reference (blurry background)

Whereas the motion signal in the foreground (moving scooter) lacks clarity of perceptual form when resolved with high temporal resolution (clear background), it appears clear and structured when detached from its immediate reference (blurry background).

In the presence of dynamic changes in the sensory image, spatio-temporally invariant information and sensitivity to change and motion have to be balanced virtually simultaneous in order to preserve visual stability and continuity. Computing perceptual structure incrementally within the persistence of the sensory signal satisfies this need to mediate between segregation of incoming information into separate representations and integration of successive input to accumulate sensory evidence over time. In this way, sensory input can be stabilized with high temporal resolution as an invariant percept, while the sensory image can still be accessed for sufficient time to update its spatio-temporal coordinates and perceive motion. An intricate balance between temporal and informational resolution is therefore crucial for perceiving the external world as both stable and continuous.

5.3 Synchrony between carrier and coding: a perceptual moment

A functional implementation of this temporal window might involve coupling between establishing a spatio-temporal reference to an external event and encoding individual information within the sensory signal. Within this coupling, encoding of perceptual structure is hardcoded within the visual analysis. The sensory signal upon which encoding is based, however, is conveyed via camera-like sampling and the latency of this carrier can vary as a function of bottom-up factors. Such a mechanism constitutes a temporal window in intermediate-level vision within which the rate of successive events determines the informational resolution of each single instance (Figure 5.4).

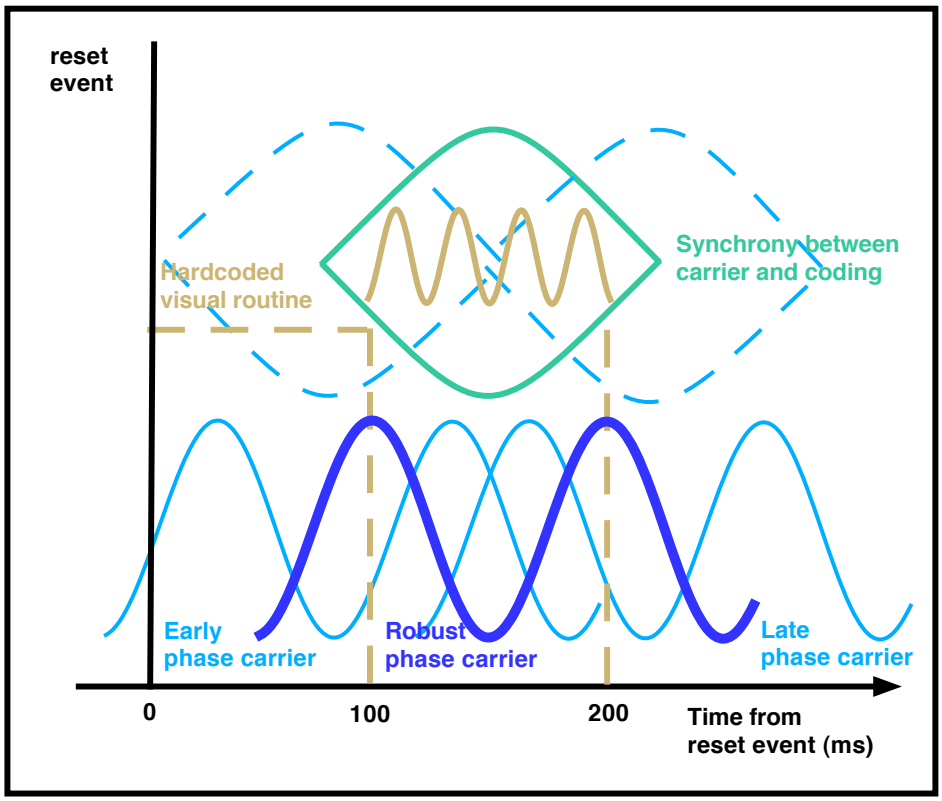


Figure 5.4 Schematic depiction how synchrony between spatio-temporal carrier and routinely encoding defines one perceptual moment of time

After a fixed interval from reset, visual routines are hardcoded within the visual analysis. Purely synchronized phase carriers deliver the sensory image out of time for critical encoding which reduces the informational capacity. Exact phase-locking between reset event and carrier, in contrast, results in perfect synchrony between access to sensory input and encoding of individual information.

Coherent perception of the external surrounding depends on successful integration of sensory information from different frames of reference. Within such reference frames, sensory signals with different latencies can be combined into a coherent percept within multisensory coordination (Mullette-Gillman, Cohen & Groh, 2005). In particular, vision has been shown to involve coding within multiple frames of reference (for review see Melcher & Colby, 2008) and the computation of perceptual form has been hypothesized to depend on the definition of a dynamic reference frame based on motion segmentation (Ögmen & Herzog, 2010).

Here, sensory information is initially encoded within a retinotopically organized frame of reference that is high in capacity and easily overwritten by subsequent input (Wundt, 1989; Sperling, 1960; Breitmeyer & Ögmen, 2006). Temporal reference to the external world is provided by precise phase-locking between reset event and cyclic frames, as rhythmic sampling of the sensory surrounding (VanRullen et al., 2007; Busch & VanRullen, 2010). Resets might either be induced top-down by saccadic eye-movements or anticipatory coding via β-oscillations, or evoked by real-world transitions. Phase-locking might be perturbed and precise temporal reference cannot be established, when successive events happen within one framing cycle. Consistent with findings from integration masking, such rapid successions of stimuli (< 100 ms) are perceived as collapsed into a single percept rather than separate events (Scheerer, 1973 a; Enns & Di Lollo, 2000; Breitmeyer & Ögmen, 2006).

Without precise temporal reference, however, the sensory signal arrives out of phase for a critical interval of encoding individual structure. Encoding of individual information is likely to be hardcoded at a fixed interval from reset as a visual routine (Ullman, 1984) within the temporal dynamics of the visual system. Individuation is a computationally complex task that involves indexing of salient items, marking previously indexed locations and multiple shifts of the processing focus (Ullman, 1984). The high level of abstraction necessary for individuation within arbitrarily complex stimulus configurations contravenes the apparent automaticity and immediateness of processing within one perceptual instant. Execution of such complex coding in real time requires the implementation of a specialized routine within the visual hierarchy (Roelfsema, Lamme & Spekreijse, 2000). Such routinely encoding is likely to be triggered after a fixed interval from reset without taking subsequent alterations in input into account.

One possible implementation of this mechanism might involve coupled networks

of nested oscillatory sub-cycles (coding individual content) within slow-wave carriers (defining the spatio-temporal context). Such neural networks are capable of representing individual information (Lisman & Idiart, 1995; Jensen et al., 2012). In support for this idea, recent evidence suggests cross-frequency interactions between α- and γ-bands in the selection of multiple visual targets (Landau & Fries, 2012; Landau, Schreyer, Van Pelt & Fries, 2013). Such cross-frequency multiplexing would greatly benefit from phase synchrony between carrier and coding (Buzsaki, 2006).

Within a stationary surrounding, precise phase alignment between the reset event (i.e. stimulus or fixation onset) and oscillatory carrier sampling can be established. In this way, the sensory signal arrives in-phase for encoding of individual information. Synchrony between carrier and coding enables read-out of information for the full time of persistence of the sensory signal. In the presence of dynamic changes in signal within one integration cycle, however, exact phase-locking between reset events and signal carrier is perturbed and the temporal reference to reset onset is fuzzy. In this case the sensory signal can only be accessed for a fraction of its actual persistence by encoding routines and individual information is lost.

In this way, the time to access the sensory signal - its effective persistence – together with phase-coupled coding determines the capacity of intermediate-level vision. Within such a temporal windowing mechanism the rate of successive events limits the informational resolution of each single instance. The need for both stable and continuous visual impressions imposes a duality between time and information in visual processing. A ‘perceptual moment’ is therefore a dynamic concept whose period varies as a function of visual information contained within it. Synchrony - between a temporal window that contains the sensory signal and encoding of individual perceptual units - instantiates one moment of time within its entire perceptual capacity.

6. References

Albrecht, A. R., & Scholl, B. J. (2010). Perceptually averaging in a continuous visual world extracting statistical summary representations over time. *Psychological Science*, *21*(4), 560-567.

Allport, D. A. (1968). Phenomenal simultaneity and the perceptual moment hypothesis. *British Journal of Psychology*, *59*(4), 395-406.

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short- term memory is set both by visual information load and by number of objects. *Psychological Science, 15*, 106–111.

Bai, O., Lin, P., Vorbach, S., Floeter, M. K., Hattori, N., & Hallett, M. (2008). A high performance sensorimotor beta rhythm-based brain-computer interface associated with human natural motor behavior. *Journal of Neural Engineering*, *5*(1), 24–35.

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron*, *76*(4), 695–711.

Baylis, G. C., & Driver, J. (1993). Visual attention and objects: evidence for hierarchical coding of location. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(3), 451.

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science, 321*, 851–854.

Blaser, E., Sperling, G., & Lu, Z. - L. (1999). Measuring the amplification of attention. *Proceedings of the national academy of science, 96*, 11681-11686.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436.

Breitmeyer, B. G., & Öğmen, H. (2006). *Visual Masking: Time slices through conscious and unconscious vision*. New York, NY: Oxford University Press.

Burr D. (1980). Motion smear. *Nature, 284*, 164–165.

Burr, D. C. (1984). Summation of target and mask metacontrast stimuli. *Perception, 13,*183–192.

Burr, D. C., Ross, J., & Morrone, M. C. (1986). Seeing objects in motion. *Proceedings of the Royal Society: Biological Sciences, 227*, 249–265.

Burr, D. C., Turi. M., & Anobile, G. (2010). Subitizing but not estimation of numerosity requires attentional resources. *Journal of Vision, 10*(6), 1-10.

Busch, N. A., Dubois, J., & Van Rullen, R. (2009). The phase of ongoing EEG oscillations predicts visual perception. *Journal of Neuroscience, 29*(24), 7869-7876.

Busch, N. A., & VanRullen, R. (2010). Spontaneous EEG oscillations reveal periodic sampling of visual attention. *Proceedings of the National Academy of Sciences*, *107*(37), 16048-16053.

Buschman, T. J., & Miller, E. K. (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science*, *315*(5820), 1860–2.

Buzsáki, G. (2006). *Rhythms of the Brain*. New York, NY: Oxford University Press.

Buzsáki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science*, *304*(5679), 1926–9.

Coltheart M. (1980). Iconic memory and visible persistence. *Perception & Psychophysics, 27,* 183–228.

Corbett, J. E., & Melcher, D. (under review). Characterizing ensemble statistics: Mean size is represented across multiple frames of reference.

Cowan, N. (2000). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87–185.

Cowan, N., Elliott, E. M., Scott Saults, J., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive psychology*, *51*(1), 42-100.

Crozier, W. J., & Wolf, E. (1941). Theory and measurement of visual mechanisms IV. Critical intensities for visual flicker, monocular and binocular.*The Journal of general physiology*, *24*(4), 505-534.

Dehaene, S. (1993). Temporal oscillations in human perception. *Psychological Science*, *4*(4), 264-270.

Dehaene, S. (2011). Conscious and Nonconscious Processes: Distinct Forms of Evidence Accumulation? *Biological Physics*, *60*, 141–168.

Dempere-Marco, L., Melcher, D., & Deco, G. (2011). A computational study of visual working memory capacity in the presence of saliency effects. *BMC Neuroscience, 12*(1), 64.

Dempere-Marco, L., Melcher, D., & Deco, G. (2012). Effective visual working memory capacity: an emergent effect from the neural dynamics in an attractor network. *Plos One, 7*(8): e42719.

Di Lollo, V. (1980). Temporal integration in visual memory. *Journal of Experimental Psychology: General, 109,* 75-97.

Dirac, P. (1958). *Principles of quantum mechanics* (4th ed.), Oxford, UK: Clarendon Press

Donner, T. H., Siegel, M., Fries, P., & Engel, A. K. (2009). Buildup of choice-predictive activity in human motor cortex during perceptual decision making. *Current Biology*, *19*(18), 1581–5.

Elder, J. H., & Goldberg, R. M. (2002). Ecological statistics of Gestalt laws for the perceptual organization of contours. *Journal of Vision*, *2*(4), 324–53.

Engel, A. K., Fries, P., & Singer, W. (2001). Dynamic predictions: oscillations and synchrony in top-down processing. *Nature reviews. Neuroscience*, *2*(10), 704–16.

Engel, A. K., & Fries, P. (2010). Beta-band oscillations--signalling the status quo? *Current Opinion in Neurobiology*, *20*(2), 156–65.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of Experimental Psychology: General*, *128*(3), 309.

Enns, J. T., & Di Lollo, V. (2000). What's new in visual masking? *Trends in Cognitive Sciences, 4*, 345–52.

Egeth, H. E., Leonard, C. J., & Palomares, M. (2008). The role of attention in subitizing: Is the magical number 1?. *Visual Cognition, 16*(4), 463-473.

Feldman, J. (2001). Bayesian contour integration. *Perception & Psychophysics*, *63*(7), 1171–1182.

Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: Evidence for a local "association field." *Vision Research*, *33*(2), 173–193.

Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in Cognitive Sciences*, *9*(10), 474–80.

Finkel, D. L., & Smythe, L. (1973). Short-term storage of spatial information. *Developmental Psychology, 9*, 424–428.

Fukuda, K., Awh, E., & Vogel, E. K. (2010). Discrete capacity limits in visual working memory. *Current Opinion in Neurobiology, 20*, 177–182.

Fukuda, K., Vogel, E., Mayr, U., & Awh, E. (2010). Quantity, not quality: The relationship between fluid intelligence and working memory capacity. *Psychonomic bulletin & review*, *17*(5), 673-679.

Folk, C. L., Egeth, H., & Kwak, H. (1988). Subitizing: Direct apprehension or serial processing? *Perception & Psychophics, 44*, 313-320.

Gegenfurtner, K. R., & Sperling, G. (1993). Information transfer in iconic memory experiments. *Journal of Experimental Psychology: Human Perception and Performance, 19*(4), 845–866.

Gho, M., & Varela, F. J. (1988). A quantitative assessment of the dependency of the visual temporal frame upon the cortical rhythm. *Journal de physiologie*, *83*(2), 95.

Gross, J., Kujala, J., Hamalainen, M., Timmermann, L., Schnitzler, A., & Salmelin, R. (2001). Dynamic imaging of coherent sources: Studying neural interactions in the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(2), 694–9.

Haber, R. N., & Standing, L. G. (1970). Direct estimates of the apparent duration of a flash. *Canadian Journal of Psychology*, *24*(4), 216.

Hamada, Y., Kado, H., & Suzuki, R. (2001). The temporal profile of interactions between sensory information from both hands in the secondary somatosensory cortex. *Clinical Neurophysiology*, *112*(7), 1326–1333.

Hamada, Y., Otsuka, S., Okamoto, T., & Suzuki, R. (2002). The profile of the recovery cycle in human primary and secondary somatosensory cortex: a magnetoencephalography study. *Clinical Neurophysiology*, *113*(11), 1787–1793.

Hamilton, W. (1859). Lectures on Metaphysics and Logic. In H.L. Mansel & J. Veitch (Eds.), *Metaphysics* (Vol. 1). Boston, MA: Gould & Lincoln.

Hanslmayr, S., Aslan, A., Staudigl, T., Klimesch, W., Herrmann, C. S., & Bäuml, K.-H. (2007) a. Prestimulus oscillations predict visual perception performance between and within subjects. *NeuroImage*, *37*(4), 1465–73.

Hanslmayr, S., Klimesch, W., Sauseng, P., Gruber, W., Doppelmayr, M., Freunberger, R., Pecherstorfer, T., & Birbaumer, N. (2007) b. Alpha phase reset contributes to the generation of ERPs. *Cerebral Cortex*, *17*(1), 1–8.

Hanslmayr, S., Gross, J., Klimesch, W., & Shapiro, K. L. (2011). The role of alpha oscillations in temporal attention. *Brain research reviews*, *67*(1), 331-343.

Hoff, H., & Pötzl, O. (1934). Über eine Zeitrafferwirkung bei homonymer linksseitiger Hemianopsie. *Zeitschrift für die gesamte Neurologie und Psychiatrie*, *151*, 599-641.

Hogben, J. H., & Di Lollo, V. (1985). Suppression of visible persistence in apparent motion. *Perception & Psychophysics*, *38*(5), 450-460.

Holtzman, J. D., & Gazzaniga, M. S. (1982). Dual task interactions due exclusively to limits in processing resources. Science, 218, 1325-1327.

Itti, L., Koch, C., & Niebur. E. (1998). A model of saliency-based visual attention for rapid scene analysis. *Institute of Electrical and Electronics Engineers Transactions on Pattern Analysis and Machine Intelligence, 20*(11), 1254–1259.

Iversen, J. R., Repp, B. H., & Patel, A. D. (2009). Top-down control of rhythm perception modulates early auditory responses. *Annals of the New York Academy of Sciences*, *1169*, 58–73.

Jensen, O., Bonnefond, M., & VanRullen, R. (2012). An oscillatory mechanism for prioritizing salient unattended stimuli. *Trends in Cognitive Sciences*, *16*(4), 200–6.

Jevons, W. S. (1871). The power of numerical discrimination. *Nature, 3,* 363–372.

Jokeit, H. (1990). Analysis of periodicities in human reaction times. *Naturwissenschaften*, *77*(6), 289-291.

Kahneman, D. (1973). *Attention and effort*. New Jersey, NY: Prentice-Hall.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: object-specific integration of information. *Cognitive Psychology, 24,* 174-219.

Kant, I. (1899). Kritik der reinen Vernunft. In R. Schmidt (ed.), *Kritik der reinen Vernunft*. Hamburg: Meiner.

Karatekin, C., & Asarnow, R. F. (1998). Working memory in childhood-onset schizophrenia and attention-deficit/hyperactivity disorder. *Psychiatry research*, *80*(2), 165-176.

Kaufman, E. L., Lord, M. W., Reese, T., & Volkmann, J. (1949). The discrimination of visual number. *American Journal of Psychology, 62,* 496–525.

Keil, J., Müller, N., Hartmann, T., & Weisz, N. (2013). Prestimulus Beta Power and Phase Synchrony Influence the Sound-Induced Flash Illusion. *Cerebral Cortex*, bhs409–.

Keil, J., Müller, N., Ihssen, N., & Weisz, N. (2012). On the variability of the McGurk effect: audiovisual integration depends on prestimulus brain states. *Cerebral Cortex*, *22*(1), 221–31.

Klein, S. E. (2001). Measuring, estimating, and understanding the psychometric function: A commentary. *Perception & Psychophysics, 63*(8), 1421-1455.

Klimesch, W., Sauseng, P., & Hanslmayr, S. (2007). EEG alpha oscillations: the inhibition-timing hypothesis. *Brain Research reviews*, *53*(1), 63–88.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology, 4,* 219–227.

Kowler, E., & Steinman, R. M. (1977). The role of small saccades in counting. *Vision Research, 17*, 141–146.

Landau, A. N., & Fries, P. (2012). Attention samples stimuli rhythmically. *Current Biology*, *22*(11), 1000–4.

Landau, A.N., Schreyer, H., Van Pelt, S., & Fries, P. (2013). Evidence for attentional sampling in the MEG gamma band response. *Journal of Vision. VSS abstract supplement.*

Lee, E. Y., Cowan, N., Vogel, E. K., Rolan, T., Valle-Inclán, F., & Hackley, S. A. (2010). Visual working memory deficits in patients with Parkinson's disease are due to both reduced storage capacity and impaired ability to filter out irrelevant information. *Brain*, *133*(9), 2677-2689.

Liljeström, M., Kujala, J., Jensen, O., & Salmelin, R. (2005). Neuromagnetic localization of rhythmic activity in the human brain: a comparison of three methods. *NeuroImage*, *25*(3), 734–45.

Lisman, J. E., & Idiart, M. A. (1995). Storage of 7 +/- 2 short-term memories in oscillatory subcycles. *Science, 267*, 1512–1515.

Loftus, G. R., Duncan, J., & Gehrig, P. (1992). On the time course of perceptual information that results from a brief visual presentation. *Journal of Experimental Psychology: Human Performance and Perception, 18,* 530-549.

Loftus, G. R., & Irwin, D. E. (1998). On the relations among different measures of visible and informational persistence. *Cognitive Psychology, 35*, 135-199.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*, 279–281.

Mach, E. (1865). Untersuchungen über den Zeitsinn des Ohres. *Sitzungsberichte der mathematisch-naturwissenschaftlichen Classe der Kaiserlichen Akademie der Wissenschaften. 51*, II. Abteilung, 133–150.

Makeig, S, Westerfield, M., Jung, T. P., Enghoff, S., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2002). Dynamic brain sources of visual evoked responses. *Science*, *295*(5555), 690–4.

Makeig, Scott, Debener, S., Onton, J., & Delorme, A. (2004). Mining event-related brain dynamics. *Trends in Cognitive Sciences*, *8*(5), 204–10.

Mandler, G., & Shebo, B. J. (1982). Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General, 111*, 1 -22.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience methods*, *164*(1), 177–90.

Marr, D. (1982). *Vision. A Computational Investigation into the Human Representation and Processing of Visual Information*. New York, NY: W. H. Freeman and Company.

Mathewson, K. E., Gratton, G., Fabiani, M., Beck, D. M., & Ro, T. (2009). To see or not to see: prestimulus alpha phase predicts visual awareness. *The Journal of Neuroscience*, *29*(9), 2725–32.

McElree, B. & Carrasco, M. (1999). The temporal dynamics of visual search: Evidence for parallel processing in feature and conjunction searches. *Journal of Experimental Psychology: Human Perception and Performance, 25*(6), 1517-1539.

Melcher, D. (2011). Visual stability. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *366*(1564), 468–75.

Melcher, D., & Colby, C. L. (2008). Trans-saccadic perception. *Trends in cognitive sciences*, *12*(12), 466-473.

Melcher, D., & Piazza, M. (2011). The role of attentional priority and saliency in determining capacity limits in enumeration and visual working memory. *PLoS One, 6*(12):e29296.

Minsky, M., & Papert, S. (1969). *Perceptrons*. Cambridge, MA: MIT Press.

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science, 229,* 782-784.

Mullette-Gillman, OD. A., Cohen, Y. E., & Groh, J. M. (2005). Eye-centered, head-centered, and complex coding of visual and auditory targets in the intraparietal sulcus. *Journal of Neurophysiology*, *94*(4), 2331-2352.

Nakayama, K., & Mackeben, M. (1989). Sustained and transient components of focal visual attention. *Vision research*, *29*(11), 1631-1647.

Neisser, U. (1967). *Cognitive psychology*. New York, NY: Appleton-Century-Crofts.

Newton, I. (1689). *Philosophiae Naturalis Principia Mathematica* I (in transl. by A. Motte (1729) *The mathematical principles of natural philosophy,* rev. F. Cajori (1934)). Berkeley, CA: University of California Press.

Nolte, G. (2003). The magnetic lead field theorem in the quasi-static approximation and its use for magnetoencephalography forward calculation in realistic volume conductors. *Physics in Medicine and Biology*, *48*(22), 3637.

Öğmen, H. (1993). A neural theory of retino-cortical dynamics. *Neural Networks, 6*, 245–273.

Öğmen, H. (2007). A theory of moving form perception: synergy between masking, perceptual grouping, and motion computation in retinotopic and non-retinotopic representations. *Advances in Cognitive Psychology, 3*, 67–84.

Öğmen, H., & Herzog, M. H. (2010). The Geometry of Visual Perception: Retinotopic and Nonretinotopic Representations in the Human Visual System. *Proceedings of the IEEE*, *98*(3), 479–492.

Okazaki, M., Kaneko, Y., Yumoto, M., & Arima, K. (2008). Perceptual change in response to a bistable picture increases neuromagnetic beta-band activities. *Neuroscience research*, *61*(3), 319–28.

Olivers, C. N. L., & Watson, D. G. (2008). Subitizing requires attention. *Visual Cognition, 16*(4), 439-462.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 1–9.

Palomares, M., & Egeth, H. (2010). How element visibility affects visual enumeration. *Vision Research, 50*(19), 2000-2007.

Palomares, M., Smith, P. R., Pitts, C. H., & Carter, B. M. (2011). The Effect of Viewing Eccentricity on Enumeration. *PLoS ONE 6*(6). e20779.

Pearson, K. (1900). On the criterion that a given system of derivations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 50*(5), 57-175.

Pelli, D. G. (1997). The VideoToolbox software for visual psycho-physics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.

Pöppel, E. (1997). A hierarchical model of temporal perception. *Trends in cognitive sciences*, *1*(2), 56-61.

Pöppel, E. (2009). Pre-semantically defined temporal windows for cognitive processing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1525), 1887-1896.

Pöppel, E., & Logothetis, N. (1986). Neuronal oscillations in the human brain. *Naturwissenschaften*, *73*(5), 267-268.

Pogosyan, A., Gaynor, L. D., Eusebio, A., & Brown, P. (2009). Boosting cortical activity at Beta-band frequencies slows movement in humans. *Current Biology*, *19*(19), 1637–41.

Phillips, W. A., & Baddeley, A. D. (1971). Reaction time and short-term visual memory. *Psychonomic Science, 22*(2), 73–74.

Piazza, M., Fumarola, A. Chinello, A., & Melcher, D. (2011). Subitizing reflects visuo-spatial object individuation capacity. *Cognition, 121*(1), 147-153.

Pylyshyn, Z. W. (1989). The role of location indexes in spatial perception: a sketch of the FINST spatial index model. *Cognition, 32,* 65-97.

Pylyshyn, Z. W. (1994). Some primitive mechanisms of spatial attention. *Cognition, 50*, 363–384.

Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition*, *80*(1), 127-158.

Pylyshyn, Z. W. (2007). *Things and places: How the mind connects with the world*. Cambridge, MA: MIT press.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spatial Vision, 3*, 179– 197.

Ramachandran V. S., Rao V. M., Vidyasagar T. R. (1974). Sharpness constancy during movement perception – short note. *Perception, 3*, 97–98.

Rayner K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372–422.

Reed, A. V. (1973). Speed-Accuracy Trade-Off in Recognition Memory. *Science, 181*(4099). 574-576.

Rensink, R. A. (2000). The dynamic representation of scenes. *Visual cognition*, *7*(1-3), 17-42.

Rieger, J. W., Braun, C., Gegenfurtner, K. R., & Bülthoff, H. (2005). The Dynamics of Visual Pattern Masking in Natural Scene processing : A MEG-Study. *Journal of Vision*, *5*(3), 275–286.

Roelfsema, P. R., Engel, A. K., König, P., & Singer, W. (1997). Visuomotor integration is associated with zero time-lag synchronization among cortical areas. *Nature*, *385*, 157–161.

Roelfsema, P. R., Lamme, V. A. F., & Spekreijse, H. (2000). The implementation of visual routines. *Vision Research, 40*, 1385-1411.

Romei, V., Brodbeck, V., Michel, C., Amedi, A., Pascual-Leone, A., & Thut, G. (2008). Spontaneous fluctuations in posterior alpha-band EEG activity reflect variability in excitability of human visual areas. *Cerebral Cortex*, *18*(9), 2010–8.

Sacks, O. (1992). *Migraine. Revised and expanded*. New York, NY: Vintage Books.

Sagi, D., & Julesz, B. (1985). Enhanced detection in the aperture of focal attention during simple discrimination tasks. *Nature, 321*, 693–695.

Sandberg, L. W. (1963). Signal distortion in nonlinear feedback systems. *The Bell system technical journal, 42*, 2533-2550.

Scheerer, E. (1973) a. Integration, interruption and processing rate in visual backward masking. *Psychologische Forschung, 36*, 71–93.

Scheerer, E. (1973) b. Integration, interruption and processing rate in visual backward masking. *Psychologische Forschung, 36*, 95-115.

Schiller, P.H. (1965). Backward masking for letters. *Perceptual & Motor Skills, 20*, 47–50.

Scholl, B. J., Pylyshyn, Z. W., & Feldman, J. (2001). What is a visual object? Evidence from target merging in multiple object tracking. *Cognition*, *80*(1), 159-177.

Shallice, T. (1964). The detection of change and the perceptual moment hypothesis. *British Journal of Statistical Psychology, 17*, 113–135.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*, 379–423.

Singer, W. (1999) a. Neuronal synchrony: a versatile code for the definition of relations? *Neuron*, *24*, 49–65.

Singer, W. (1999) b. Time as coding space?. *Current Opinion in Neurobiology, 9*(52), 189-194.

Smith, B.C. (1998). *On the origin of objects*. Cambridge, MA: MIT Press

Spelke, E. (1988). Where perceiving ends and thinking begins: the apprehension of objects in infancy. In A. Yonas (Ed.), *Perceptual development in infancy* (pp. 197-234). Hillsdale, NJ: Erlbaum.

Sperling, G. (1960). The information available in brief visual presentations. *Psycholological Monographs: General and Applied, 74(11)*, 1–29.

Sperling, G. (1963). A model for visual memory tasks. *Human Factors, 5*, 19-31.

Stroud, J.M. (1955). The fine structure of psychological time. In H. Quastler (Ed.), *Information theory in psychology: Problems and methods* (pp. 174–207). Glencoe, IL: Free Press.

Tallon-Baudry, C., Bertrand, O., Delpuech, C., & Pernier, J. (1996). Stimulus Specificity of Phase-Locked and Non-Phase-Locked 40 Hz Visual Responses in Human. *The Journal of Neuroscience*, *16*(13), 4240–4249.

Thut, G., Nietzel, A., Brandt, S. A., & Pascual-Leone, A. (2006). Alpha-band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection. *The Journal of Neuroscience*, *26*(37).

Tipper, S. P., Brehaut, J. C., & Driver, J. (1990). Selection of moving and static objects for the control of spatially directed action. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(3), 492.

Townsend, J. T. (1971). A note on the identifiability of parallel and serial processes. *Perception & Psychophysics, 10*(3), 161-163.

Treisman, A. & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology, 12*, 97-136.

Ullman, S. (1984). Visual routines. *Cognition, 18,* 97-159.

Van Dijk, H., Schoffelen, J.-M., Oostenveld, R., & Jensen, O. (2008). Prestimulus oscillatory activity in the alpha band predicts visual discrimination ability. *The Journal of Neuroscience*, *28*(8), 1816–23.

VanRullen, R. (2011). Four common conceptual fallacies in mapping the time course of recognition. *Frontiers in Psychology*, *2*, 365.

VanRullen, R., & Thorpe, S. J. (2001). Rate Coding Versus Temporal Order Coding: What the Retinal Ganglion Cells Tell the Visual Cortex. *Neural Computation*, *13*(6), 1255–1283.

VanRullen, R., & Koch, C. (2003). Is perception discrete or continuous?. *Trends in Cognitive Science, 7*(5), 207-213.

VanRullen, R., Carlson, T., & Cavanagh, P. (2007). The blinking spotlight of attention. *Proceedings of the National Academy of Sciences*, *104*(49), 19204-19209.

Van Veen, B. D., Van Drongelen, W., Yuchtman, M., & Suzuki, A. (1997). Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Trans. Biomed. Eng*, *44*(9), 867–880.

Varela, F. J., Toro, A., Roy John, E., & Schwartz, E. L. (1981). Perceptual framing and cortical alpha rhythm. *Neuropsychologia*, *19*(5), 675–686.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 32*(6), 1436-1451.

Volberg, G., Wutz, A., & Greenlee, M. W. (2013). Top-down control in contour grouping. *PloS one*, *8*(1), e54085.

von Baer, K. E. (1864). Welche Auffassung der lebenden Natur ist die richtige? Und wie ist diese Auffassung auf die Entomologie anzuwenden? In H. Schmitzdorff (ed.) *Reden gehalten in wissenschaftlichen Versammlungen und kleinere Aufsätze vermischten Inhalts* (pp. 237–284). St. Petersburg, Russia: Verlag der kaiserlichen Hofbuchhandlung.

Von der Malsburg, C., & Schneider, W. (1986). A neural cocktail-party processor. *Biological Cybernetics*, *54*(1), 29–40.

Waldert, S., Preissl, H., Demandt, E., Braun, C., Birbaumer, N., Aertsen, A., & Mehring, C. (2008). Hand movement direction decoded from MEG and EEG. *The Journal of Neuroscience*, *28*(4), 1000–8.

Watson, D., G., Maylor, E., A., & Bruce, L., A., M. (2007). The role of eye movements in subitizing and counting. *Journal of Experimental Psychology: Human Perception and Performance, 33*(6), 1389-1399.

Watson, D., G., Maylor, E., A., Allen, G. E. J., & Bruce, L., A., M. (2007).Early visual tagging: Effects of target-distractor similarity and old age on search, subitization, and counting. *Journal of Experimental Psychology: Human Perception and Performance, 33*(3), 549-569.

Wertheimer, M. (1912). Experimentelle Studien über das Sehen von Bewegung. Zeitschrift für Psychologie, 61,161–265.

Wertheimer, M. (1923). Lehre von der Gestalt II. *Psychologische Forschung*, *4*, 301–350.

Wikström, H., Huttunen, J., Korvenoja, A., Virtanen, J., Salonen, O., Aronen, H., & Ilmoniemi, R. J. (1996). Effects of interstimulus interval on somatosensory evoked magnetic fields (SEFs): a hypothesis concerning SEF generation at the primary sensorimotor cortex. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *100*(6), 479–487.

Winkler, I., Reinikainen, K., & Näätänen, R. (1993). Event-related brain potentials reflect traces of echoic memory in humans. *Perception & Psychophysics*, *53*(4), 443–9.

White, C. T. (1963). Temporal numerosity and the psychological unit of duration. *Psychological Monographs: General and Applied*, *77*(12), 1-37.

Wühle, A., Mertiens, L., Rüter, J., Ostwald, D., & Braun, C. (2010). Cortical processing of near-threshold tactile stimuli: An MEG study. *Psychophysiology*, *47*(3), 523–534.

Wundt, W. (1899). Zur Kritik tachistoskopischer Versuche. *Philosophische Studien, 15,* 287-317.

Wundt, W. (1900). Zur Kritik tachistoskopischer Versuche II. *Philosophische Studien, 16*, 61–81.

Wutz, A., Caramazza, A., & Melcher, D. (2012). Rapid enumeration within a fraction of a single glance: The role of visible persistence in object individuation capacity. *Visual Cognition, 20*(6), 717-732.

Wutz, A., & Melcher, D. (2013). Temporal buffering and visual capacity: The time course of object formation underlies capacity limits in visual cognition. *Attention, Perception, & Psychophysics*, *75*(5), 921–933.

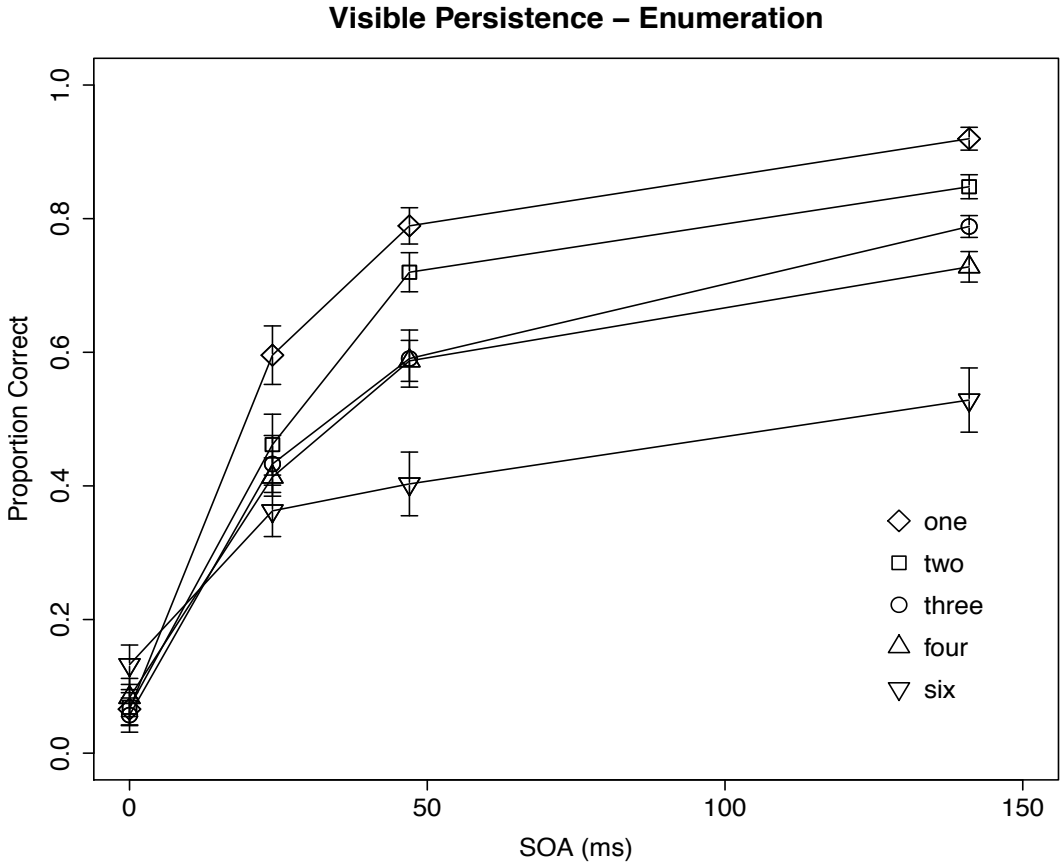
Xu, Y., & Chun, M. M. (2009). Selecting and perceiving multiple visual objects. *Trends in Cognitive Sciences, 13*(4), 167–174.

Zylberberg, A., Dehaene, S., Roelfsema, P. R., & Sigman, M. (2011). The human Turing machine: a neural framework for mental programs. *Trends in Cognitive Sciences*, *15*(7), 293–300.

Supplementary Material

Chapter 2

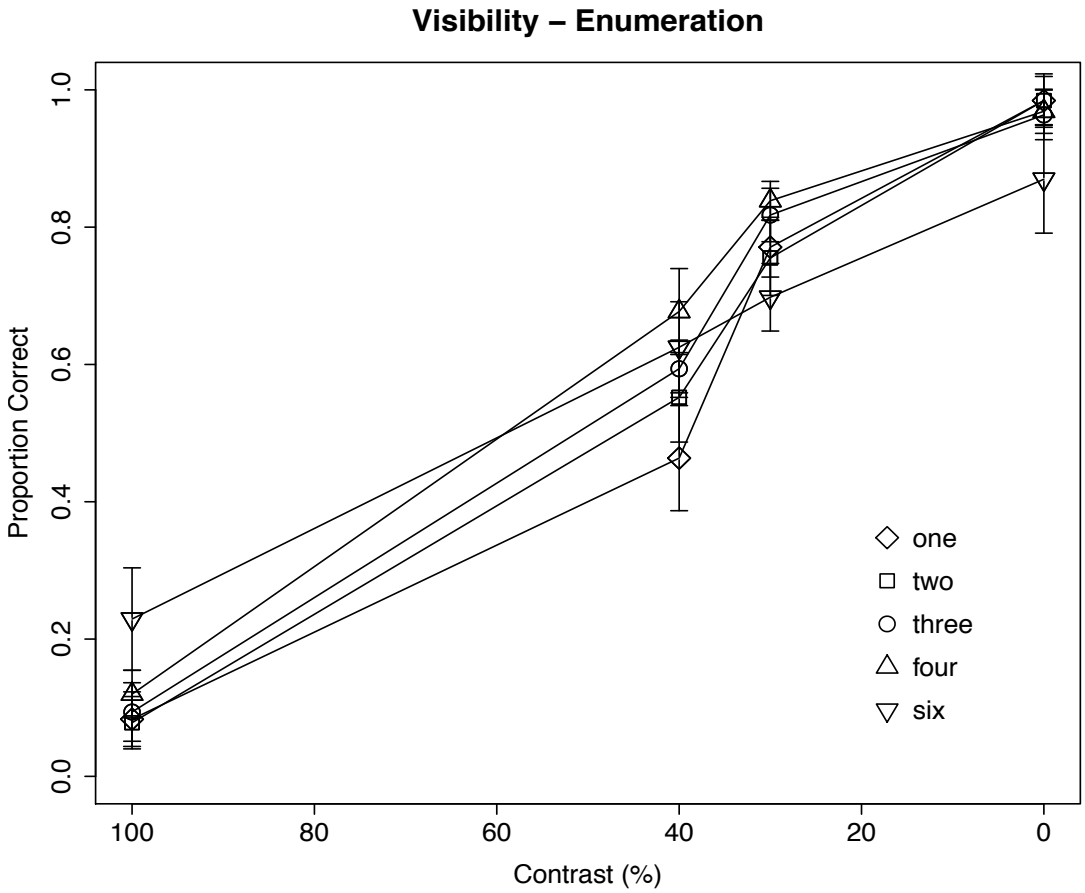
SF2.1



Results of the enumeration experiment with the manipulation of SOA

Proportion of correct trials as a function of SOA for different item numerosities. Error bars display one standard error of the mean for within-subject designs. Individual performance values have been centered on the mean performance of each subject before calculating the standard error.

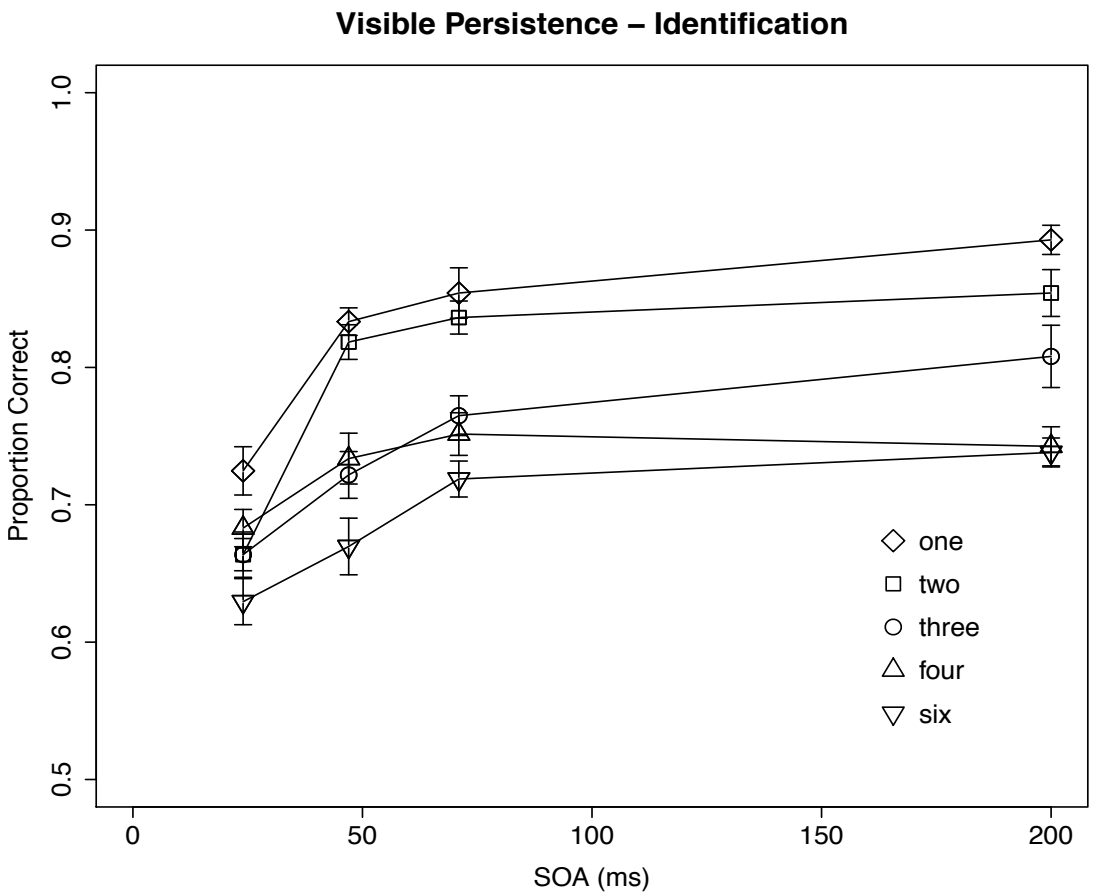
SF2.2



Results of the enumeration experiment with the manipulation of mask contrast

Proportion of correct trials as a function of mask contrast for different item numerosities. Error bars display one standard error of the mean for within-subject designs. Individual performance values have been centered on the mean performance of each subject before calculating the standard error.

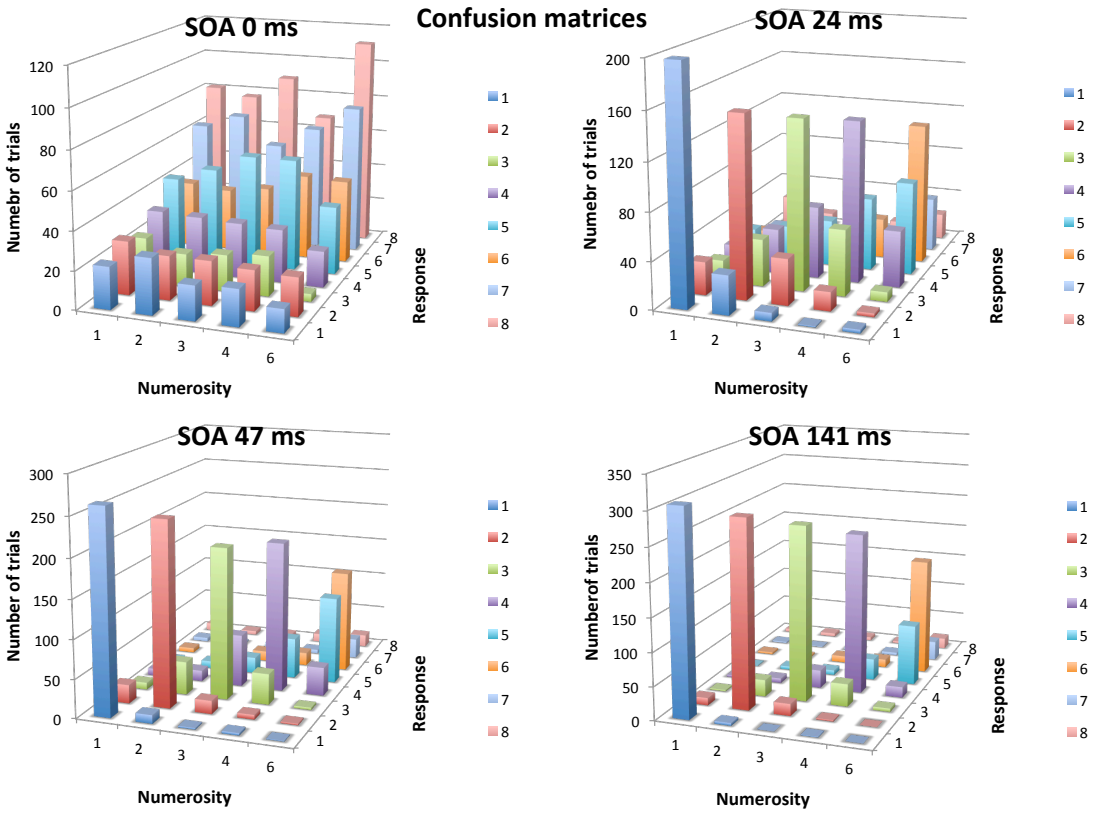
SF2.3



Results of the identification experiment with the manipulation of SOA

Proportion of correct trials as a function of SOA for different item set sizes. Error bars display one standard error of the mean for within-subject designs. Individual performance values have been centered on the mean performance of each subject before calculating the standard error.

SF2.4



Response matrices displaying the number of responses across all trials and subjects for each item numerosity and mask-item(s) SOA for the enumeration experiment with the manipulation of SOA

ST2.1 Results of the two-way within-subjects ANOVA with SOA (0, 24, 47 & 141 ms) and Numerosity (1,2,3) as factors for the enumeration experiment with the manipulation of SOA

Factor	*df1*	*df2*	*F*	*p* <	η_p^2
SOA	3	39	176.8	.001	.932
Numerosity	2	26	26.6	.001	.671
SOA x Numerosity	6	78	5.0	.001	.279

ST2.2 Descriptive statistics of the enumeration experiment with the manipulation of SOA

Mean and standard deviation of accuracy and reaction time measures for each item numerosity and SOA

	M (accuracy)	*SD (accuracy)*	*M (RT in s)*	*SD (RT in s)*
SOA0_1	0.07	0.08	1.36	0.28
SOA0_2	0.07	0.07	1.40	0.29
SOA0_3	0.06	0.06	1.48	0.35
SOA0_4	0.08	0.07	1.40	0.27
SOA0_6	0.13	0.08	1.50	0.32
SOA24_1	0.60	0.21	1.13	0.23
SOA24_2	0.46	0.21	1.18	0.24
SOA24_3	0.43	0.18	1.26	0.24
SOA24_4	0.41	0.13	1.38	0.28
SOA24_6	0.36	0.15	1.63	0.39
SOA47_1	0.79	0.16	0.98	0.15
SOA47_2	0.72	0.17	1.05	0.18
SOA47_3	0.59	0.21	1.23	0.26
SOA47_4	0.59	0.15	1.39	0.33
SOA47_6	0.40	0.19	1.66	0.41
SOA140_1	0.92	0.10	0.90	0.14
SOA140_2	0.85	0.10	0.97	0.16
SOA140_3	0.79	0.12	1.06	0.17
SOA140_4	0.73	0.11	1.30	0.28
SOA140_6	0.53	0.19	1.75	0.46

ST2.3 Descriptive statistics of the enumeration experiment with the manipulation of mask contrast

Mean and standard deviation of accuracy and reaction time measures for each item numerosity and mask contrast

	M (accuracy)	*SD (accuracy)*	*M (RT in s)*	*SD (RT in s)*
100%_1	0.08	0.14	1.34	0.45
100%_2	0.08	0.13	1.22	0.34
100%_3	0.09	0.13	1.26	0.34
100%_4	0.12	0.10	1.28	0.36
100%_6	0.23	0.16	1.31	0.37
40%_1	0.46	0.27	1.20	0.39
40%_2	0.55	0.25	1.12	0.36
40%_3	0.59	0.18	1.16	0.39
40%_4	0.68	0.26	1.26	0.44
40%_6	0.63	0.21	1.58	0.50
30%_1	0.77	0.20	0.97	0.35
30%_2	0.76	0.21	0.98	0.33
30%_3	0.82	0.19	1.03	0.30
30%_4	0.84	0.15	1.14	0.34
30%_6	0.70	0.16	1.61	0.48
0%_1	0.98	0.04	0.68	0.14
0%_2	0.98	0.04	0.72	0.11
0%_3	0.96	0.06	0.81	0.15
0%_4	0.97	0.04	0.96	0.26
0%_6	0.87	0.19	1.51	0.54

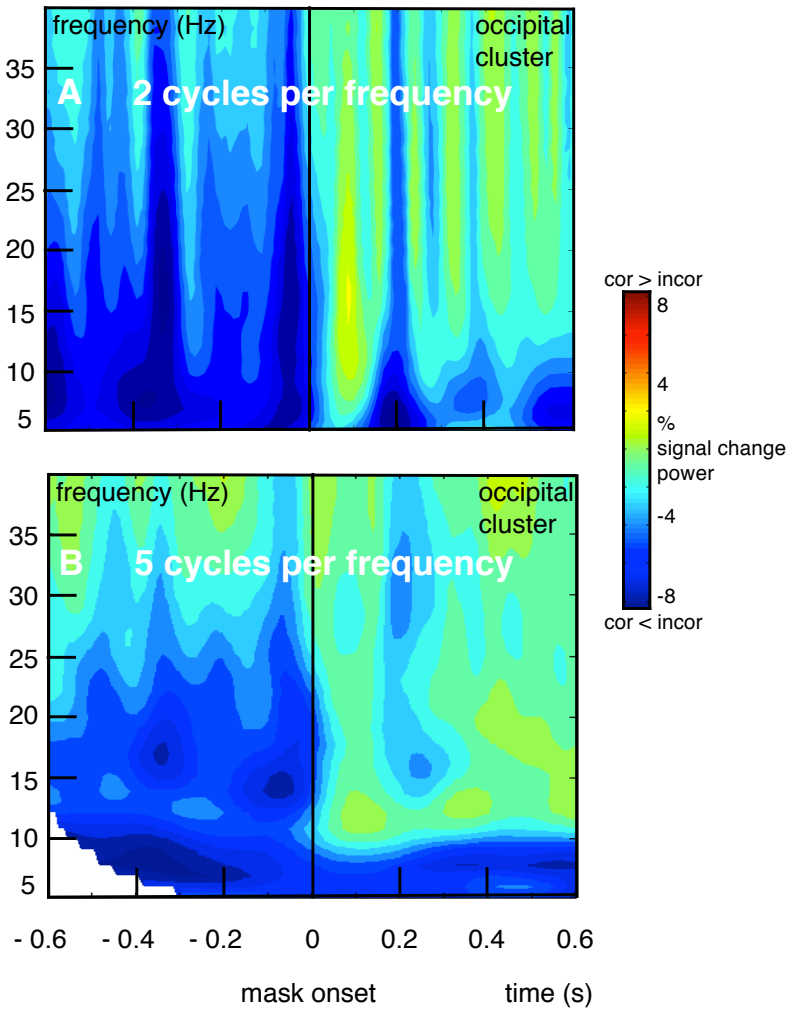
ST2.4 Descriptive statistics of the identification experiment with the manipulation of SOA

Mean and standard deviation of accuracy and reaction time measures for each item set size and SOA

	M (accuracy)	*SD* (accuracy)	*M* (RT in s)	*SD* (RT in s)
SOA24_1	0.72	0.09	0.91	0.23
SOA24_2	0.66	0.10	0.91	0.19
SOA24_3	0.66	0.08	0.93	0.23
SOA24_4	0.68	0.08	0.95	0.20
SOA24_6	0.63	0.10	0.95	0.23
SOA47_1	0.83	0.08	0.83	0.22
SOA47_2	0.82	0.06	0.88	0.20
SOA47_3	0.72	0.09	0.91	0.27
SOA47_4	0.73	0.12	0.98	0.33
SOA47_6	0.67	0.09	0.98	0.23
SOA71_1	0.85	0.13	0.79	0.19
SOA71_2	0.84	0.09	0.85	0.20
SOA71_3	0.76	0.11	0.85	0.18
SOA71_4	0.75	0.09	0.92	0.18
SOA71_6	0.72	0.07	0.95	0.20
SOA200_1	0.89	0.09	0.76	0.20
SOA200_2	0.85	0.09	0.80	0.18
SOA200_3	0.81	0.12	0.84	0.18
SOA200_4	0.74	0.08	0.87	0.16
SOA200_6	0.74	0.07	0.97	0.30

Chapter 4

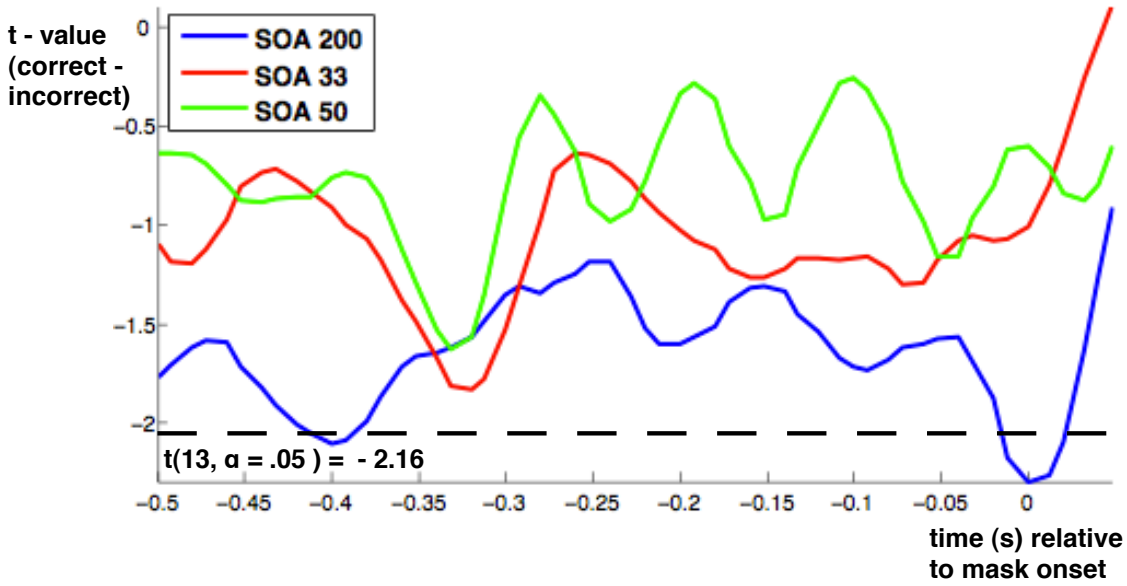
SF4.1



Time-frequency plot showing the percent in signal change in oscillatory power within correct in comparison to incorrect trials ((correct-incorrect)/incorrect) around mask onset within an occipital cluster of sensors shown in Figure **4.2.B** & **C**

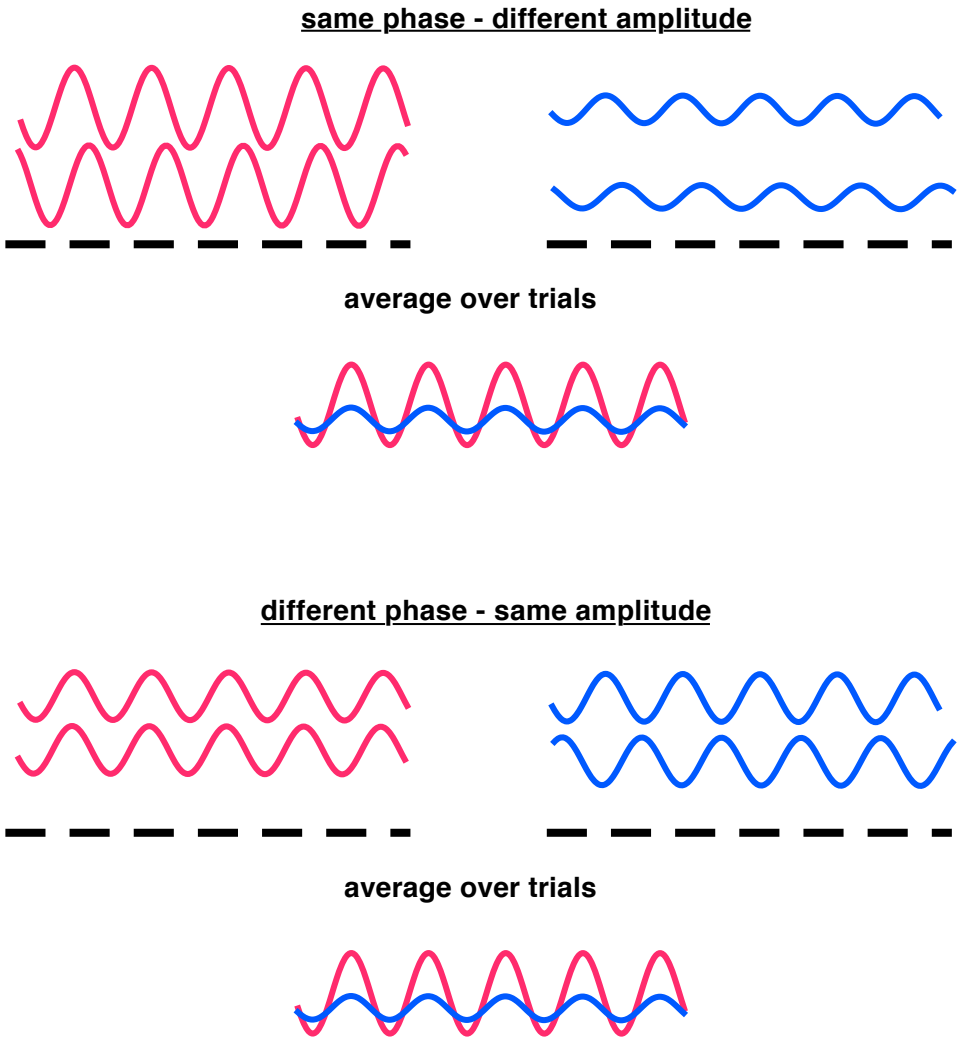
Panel A: time-frequency representation with 2 cycles per frequency; Panel B: time-frequency representation with 5 cycles per frequency. Both figures, but panel A with a better temporal resolution, show higher oscillatory power between 15 and 20 Hz for incorrect compared to correct trials solely in intervals prior to mask onset. Both foci of effects around -350 ms and -50 ms before mask onset are present in both figures, but in B with a better frequency resolution, and well separable from post-stimulus activity.

SF4.2



Temporal evolution of t-values (df=13) of the contrast correct vs. incorrect trials at 15 Hz for the 3 different SOAs (33 ms (red), 50 ms (green) and 200 ms (blue)) averaged over the occipital cluster of sensor locations depicted in Figure **4.2 B & C**. Only within 200 ms SOA trials the effect is large enough to cross the critical significance threshold ($\alpha < .05$).

SF4.3



Schematic depiction of possible reasons for differences in event-related fields between two conditions

In the upper panel, higher amplitude in one condition (red) yields also higher amplitude in the evoked field averaged over trials. In the lower panel, both conditions are equal in amplitude, but the more consistent phase concentration in one condition (red) yields higher amplitude in the evoked field averaged over trials.

ST4.1 Average number of trials over subjects (N=14) within the respective condition (120 trials in total per SOA x set size cell)

SOA 30

	response					
set size	**no resp**	***0***	**1**	**2**	**3**	**4**
1	1.4286	*12.929*	**92.857**	7.9286	3.9286	0.92857
2	1.6429	*7.7857*	15.643	**76.786**	15.286	2.8571
3	2.8571	*7.9286*	7.5714	22.786	**67.857**	11

SOA 50

	response					
set size	**no resp**	***0***	**1**	**2**	**3**	**4**
1	1.0714	*7.8571*	**104.07**	4.7143	1.7857	0.5
2	1.0714	*6*	10.929	**89.857**	10.429	1.7143
3	1.9286	*5*	4.3571	17.929	**81.571**	9.2143

SOA 200

	response					
set size	**no resp**	***0***	**1**	**2**	**3**	**4**
1	0.7857	*3.9286*	**111.29**	2.7857	1.2143	0
2	1.5714	*2.1429*	6.9286	**105.64**	3.4286	0.28571
3	1.3571	*2.2143*	3.2143	8.4286	**100.79**	4
4	3.2143	*2.2143*	2.3571	1.7143	25.286	**85.143**

The correct response is highlighted in bold and the most common error shown in red. The most frequent error across all conditions is one item less than actually presented. The column labeled “0” depicts the number of “no detection” trials (*in italics*) in which observers missed the onset of the target display entirely.