THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
2011, 64 (4), 767–791

Psychology Press
Taylor & Francis Group

Is there brief temporal buffering of successive visual inputs?

Wayne S. Smith
Durham University, Durham, UK

J. D. Mollon
Department of Experimental Psychology, University of Cambridge, Cambridge, UK

Rishi Bhardwaj and Hannah E. Smithson
Durham University, Durham, UK

The part-report advantage has been used to identify short-lived forms of visual storage (Sperling, 1960). We adopt the part-report paradigm to test whether visual memory can preserve, for a brief time, successive inputs and their temporal order. In our experiments, two successive arrays, each of 4 digits, were presented on each trial. The two arrays were spatially coincident, and each was followed by a random pattern-mask. In the part-report conditions, an auditory cue indicated whether the participant should report the first array or the second array. The results consistently showed a part-report advantage, which ranged in size from 16% to 37%. Delaying the cue by 500 ms abolished most of this advantage, in that performance was then similar to that in whole-report conditions. Subsequent experiments confirmed that the part-report superiority we measure is not achieved by (a) making eye movements that spatially displace the second array relative to the first; (b) extracting information from a single snapshot containing an integrated representation of the targets and masks; or (c) transferring a subset of material to a phonological store. We propose instead that observers have access to a limited, rapidly decaying representation of successive visual inputs stored in temporal sequence.

Keywords: Iconic memory; Visual short-term memory; Part-report; Pattern masking; Temporal storage.

To interact successfully with the physical and social world, human beings must be able to store brief mental representations of spatio-temporal structures and not merely static images. In traditional models however, visual sensory storage corresponds to a single snapshot of the world, a snapshot that is overwritten by any after-coming stimulus (Gegenfurtner & Sperling, 1993; Keysers, Xiao, Földiák, & Perrett, 2005; Tatler, 2001). We here ask instead whether successive visual inputs can be held in a buffer that preserves their temporal order.

Correspondence should be addressed to Wayne S. Smith, Centre for Software Reliability, 11th Floor, Claremont Tower, School of Computing Science, University of Newcastle, Newcastle upon Tyne, NE1 7RU, UK. E-mail: waynestuartsmith@hotmail.com
We thank our reviewers for their helpful comments on a previous version of the manuscript. Thanks are also due to all the participants.

http://www.psypress.com/qjep
DOI:10.1080/17470218.2010.511237

Many studies of human vision have shown the existence of iconic storage, a high-capacity fragile sensory memory that decays within a few hundred milliseconds (Averbach & Coriell, 1961; Chow, 1986; Coltheart, 1983; Neisser, 1967; Palmer, 1988; Sperling, 1960; Treisman, Russell, & Green, 1975). Coltheart (1980) made a classical distinction between "visual persistence" and the "informational persistence" that underlies iconic storage. The former exhibits an inverse relationship to stimulus intensity, which can be explained by the shortening of photoreceptor time constants in light adaptation (Stockman, Langendörfer, Smithson, & Sharpe, 2006). "Informational persistence" is less affected by physical parameters of the stimulus. It is plausible to suppose that informational persistence underlying iconic storage occurs concurrently at several levels of visual analysis, some of which would conventionally be considered precategorical, some postcategorical.

Iconic storage and "fragile visual short-term memory"

Iconic storage has been considered to be quite distinct from visual short-term memory (VSTM), which has a smaller capacity but is more robust and lasts for many seconds (Phillips, 1974). However, it is becoming increasingly difficult to distinguish iconic storage from a fragile form of VSTM (Sligte, Scholte, & Lamme, 2008). Two experimental properties previously used to define iconic storage—a part-report superiority effect and susceptibility to an after-coming pattern mask—turn out to characterize also the longer lasting representations that have been identified as VSTM.

Classically the high-capacity, rapidly decaying iconic store is operationally revealed by the part-report paradigm: When a postcue instructs a participant to report only a randomly chosen subset of the presented items, then he or she appears to have available many more items than can be given back in a whole-report condition. The argument is that the participant is able to use the cue to retrieve the specified items from a rapidly decaying trace and transfer them to a more durable store from which they can be reported (Neisser, 1967; Sperling, 1960, 1963). However, part-report effects have also been found under experimental conditions that are thought to tap VSTM (Griffin & Nobre, 2003; Landman, Spekreijse, & Lamme, 2003; Makovski & Jiang, 2007) or a fragile form of VSTM (Sligte et al., 2008). For example, in the experiment of Sligte, Scholte, and Lamme, a 32-element array was displayed for 250 ms and followed after some interval by a probe array. The participant's task was to decide whether a cued element had changed between the two arrays. The advantage conferred by the cue was obtained when the cue was delayed by intervals of 1,000 ms or more, intervals that are held to characterize VSTM.

A second property that defined the classical iconic store is that it is displaced or overwritten by an after-coming pattern mask. However, similar effects have been found for VSTM and have further blurred the distinction between iconic storage and VSTM. For example, in the experiments of Makovski and Jiang (2007) and Sligte et al. (2008), an after-coming array of irrelevant items disrupted memory for a target array presented earlier. In the case of both iconic storage and VSTM the after-coming stimulus may impair performance either by occupying the same visual analysers as the target or by deflecting attention in space or time.

A buffer for the storage of sequences?

Visual perception must operate in a dynamic and changing world where important information frequently changes over time (Freyd, 1987; Robbins, 2004). Current models of VSTM acknowledge that a sequence of inputs may be stored, provided that they are separated by intervals of the order of 500 ms or more (Jiang & Kumar, 2004). In contrast, the traditional and long-standing view of iconic or visual sensory memory is that single snapshots of the world are stored for short periods of time whilst they are made available to short-term memory. Each successive input is overwritten by the next (Keysers et al., 2005), as when, for example, the eye makes a saccade (Tatler, 2001).

And indeed this has been put forward as an explanation of change blindness (Becker, Pashler, & Anstis, 2000; Landman et al., 2003). The fragile form of VSTM recently proposed by Sligte et al. (2008) has, like iconic memory, been envisaged as holding single snapshots.

In this paper we wish to raise the possibility that there is an early perceptual buffer that can store dynamic visual information, such as sequences, actions, and trajectories. That is to say, there exists a visual buffer that is four-dimensional, encoding time as well as the spatial dimensions. This buffer would have many properties in common with both iconic storage and the fragile form of VSTM. The possibility that iconic memory may store items that are dispersed both in space and in time has been raised previously (Schill & Zetzsche, 1995), and some empirical evidence is offered by the study of Smithson and Mollon (2006), which dissociated the two experimental operations used to define the icon—the part-report and the masking paradigms. A target array of letters was followed after 100 ms by a high-contrast chequerboard mask. A part-report cue following the mask still supported a part-report advantage, suggesting that independent representations of target and mask were concurrently held in a short-lived visual buffer.

In the case of the *auditory* system, it has often been suggested that there is a brief sensory storage of acoustic or phonemic features in their temporal sequence, a storage that allows the recognition of repeated sequences of white noise (Guttman & Julesz, 1963) and the retrospective reparsing of phonemic sequences such as "ice cream and apple pie" and "I scream and I yell". This *echoic memory* (Neisser, 1967) is intrinsically temporal in that it holds phonemic sequences. In the classical experiment by Darwin, Turvey, and Crowder (1972), participants were presented with three streams of spoken letter sequences at three different locations and were then cued to report items from one of the three locations, chosen at random. As in studies of iconic storage, Darwin et al. found a part-report advantage: The number of items available in part report, estimated by multiplying by three the number correct per stream, was greater than the total number of items that could be reported when recovery of all three streams was required. The size of the advantage—16%—was smaller than that found by Sperling (1960) for visual arrays of letters.

Pattern masks are routinely used in the fields of visual cognition and psycholinguistics to "terminate processing" of an earlier target. In the traditional "snapshot" account of very-short-term visual storage, an after-coming high-contrast mask acts either by displacing the target from the representation ("interruption") or by being superposed on the representation ("integration"). In either case, if there does exist a short-lived visual buffer that can hold successive inputs separately in sequence, in the way that successive phonemes are held in an echoic store, then it becomes invalid to rely on an after-coming mask to limit the time for which a target is available for further processing (Smithson & Mollon, 2006).

The present study

The part-report paradigm, originally used by Sperling (1960) to establish the existence of iconic memory and by Darwin et al. (1972) to establish its auditory analogue, was used in the present study to measure the storage and retrieval of serially presented visual information.

In a traditional experiment on iconic storage, a matrix of characters is displayed for a short duration, and then, after removal of the visual stimulus, a cue instructs the participant to report a randomly chosen row of characters from that matrix. In the present study, we adopt the classic part-report paradigm but now the arrays are presented sequentially in time rather than simultaneously within a single frame, and the cue requires selection on the basis of temporal position rather than spatial position. By asking whether participants can select on the basis of temporal position, we test whether an ordered sequence of inputs can be held within a perceptual buffer. Two successive arrays, each of four digits, were presented on every trial. The two arrays were spatially coincident, and each was followed by a random pattern-mask. In the part-report conditions, an auditory cue indicated whether the participant should report the first array or the second array. We asked whether the participant could retrospectively select on the basis of a cue that identified a temporal subset of the stimuli.

EXPERIMENT 1

In this experiment there were three main conditions: a whole-report condition, in which participants were asked to report as many digits as possible from the two successive target arrays; a part-report no-delay condition, in which a cue immediately after the end of the stimulus sequence indicated which of the two successive target arrays to report; and a part-report delayed-cue condition, in which the cue was presented 500 ms after the end of the stimulus sequence. In addition, a subset of participants completed two precue blocks of trials in which they were informed at the start of the block to attend only to one of the two target arrays.

Method

Participants
There were 12 participants: 10 naïve and 2 practised (authors W.S.S. and H.E.S.). A subset of 8 participants completed additional "precue" conditions.

Apparatus
Visual stimuli were presented on a 20-inch (0.5-m) Mitsubishi Diamond Pro 2070 SB monitor (displaying 1,024 × 768 pixels at a frame rate of 140 Hz, giving a frame-to-frame interval of 7.14 ms). Graphics output was controlled by a Cambridge Research Systems (CRS) ViSaGe unit and a 4 × PCI express graphics card housed in an Evesham Intel® Core™2 Duo 2 × 2.13-GHz CPU computer. The experiment was executed and controlled using the Matlab® programming language (Version 6.5). The display was gamma corrected (linearized) from measurements made with a CRS OptiCAL.

Auditory stimuli were generated by a Realtek onboard soundcard with Labtec speakers. Physical synchronization of the auditory and visual stimuli was checked empirically using an oscilloscope to compare the signal from the soundcard and the output from a fast photodiode.

Stimuli
The background luminance of the stimulus monitor was 37.7 cd/m^2. Target stimuli were 1-by-4 arrays of digits. They were displayed at a luminance of 74.7 cd/m^2 and rendered in the freely available 7-segment font, "DS-Digital". The height of each letter was fixed at 29 pixels (or 10.5 mm) and subtended 0.46 degrees of visual angle at the viewing distance of 1.3 m. Each of the 4 digits in a target array was centred within the elements of an invisible grid such that the centre spacing between digits was fixed at 0.7 deg.

The use of a 7-segment font, and the controlled spacing between digits, meant that digit segments in the second array would spatially overlay corresponding digit segments from the first array. Each numeric array was constructed by selecting 4 digits at random from the set of 0 to 9 digits with replacement. Performance level for correctly guessing a digit from an array was therefore 1/10.

Mask stimuli were random 1-bit chequerboards subtending 5.6 × 1.6 degrees of visual angle. They therefore extended beyond the digit arrays on all sides. Half of the checks within the mask were set to the background level of 37.7 cd/m^2 and the remainder to the maximum luminance of the display monitor, 74.7 cd/m^2 (the same value as that used to generate the target digits). Checks within the mask were sized 4 × 4 pixels, the stroke-width of digits in the arrays.

The sequence of visual stimuli was the same for all trials and all conditions (see Figure 1 for a schematic of stimulus presentation). Since the display monitor was running at a frame-rate of 140 Hz, the frame-to-frame interval was 7.14 ms (for conversion to effective stimulus durations see Bridgeman, 1998). Each target array was presented for 3 frames followed by a uniform field for 15 frames and then a chequerboard mask for

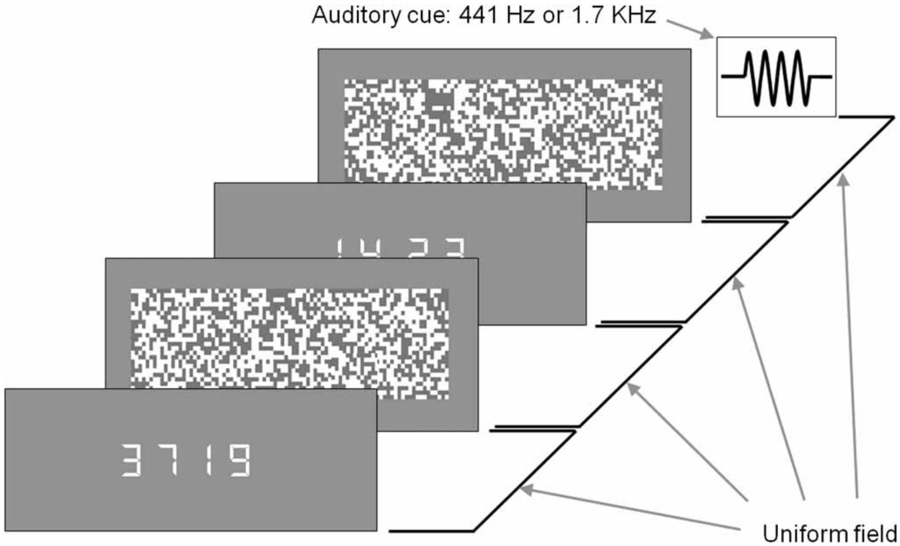

Figure 1. *Schematic representation of the stimuli for the part- and whole-report tasks in Experiment 1. Two target arrays comprising 4 digits were presented, each for 3 consecutive frames (at a frame rate of 140 Hz, giving a nominal duration of 21 ms). Each was followed by a uniform field for 15 frames (~107 ms) and then a binary noise mask for 3 frames (~21 ms). A uniform field displayed for 15 frames was also inserted between the first mask and second target array. For the part-report conditions, an auditory cue was presented after the second mask (with a delay of either 0 or 500 ms) to instruct participants to report verbally the numbers from either the first or the second array.*

3 frames. Two such sequences were presented, separated by a uniform field for 15 frames. So, the sequence was: Target Array 1 (21)−blank field (107)−mask (21)−blank field (107)−Target Array 2 (21)−blank field (107)−mask (21)−blank, where the numbers in parentheses represent the nominal duration of each item in milliseconds. The interstimulus interval (ISI) between the two target arrays was thus 235 ms.

In the part-report no-delay condition an auditory part-report cue immediately followed the offset of the second mask (i.e., 384 ms after the offset of the first digit-array and 128 ms after the offset of the second digit-array). In the part-report delayed-cue condition, the auditory cue was delayed by 500 ms after the offset of the second mask (i.e., 884 ms after the offset of the first digit-array and 628 ms after the offset of the second digit-array). The auditory cue was either a high tone (1.7 kHz) instructing participants to report the digits from the first array or a lower tone (441 Hz) instructing participants to report the digits from the second array. For the whole-report condition, an auditory tone of 441 Hz was presented after the offset of the second mask, instructing participants to report all the eight digits. The 441 Hz tone was also presented three times before the onset of the stimuli, with a 1-second separation between each of the tones. These consecutive tones provided a countdown to the onset of the stimuli and also provided a reference tone to compare to the following auditory cue (i.e., the cue tone was either higher than or the same as the prestimulus tone). All auditory tones were presented for a duration of 20 ms. A fixation cross appeared in the centre of the display screen for the duration of the reference tones.

Procedure

Each participant obtained data in separate blocks for whole-report, part-report no-delay, and part-report delayed-cue conditions, with 32 trials per block. The order of conditions was randomized for each participant. For part report, each block contained 16 trials with cues to the first array and 16 trials with cues to the second array. The cueing of first or second array was randomized within a block.

In addition, 8 of the 12 participants completed one block of 16 precue trials, in which they were asked to report digits from the first array, and a second block of 16 precue trials, in which they were asked to report digits from the second array: These conditions allowed us to test the relative legibility of the two target arrays. For the subset of 8 participants the precue blocks were randomly interleaved with the three main blocks.

Participants were instructed to report the target digits in their correct spatial and temporal positions and to guess any digits about which they were unsure. Participants gave their responses orally on hearing the tone that followed the stimulus sequence. The experimenter (author W.S.S.) entered the participant's responses into the computer via the keyboard. Participants were allowed to practise for a few minutes at the start of the experiments to familiarize themselves with the task and the cue tones.

Results

Performance in the precue blocks, in which participants were instructed in advance to report only the first array or only the second array, provides a baseline measure of the legibility of each target array when embedded in the experimental sequence. Mean scores were 85.0% (i.e., 3.40 digits out of a possible 4) for the first array and 84.4% (i.e., 3.38 digits out of a possible 4) for the second. A paired t test comparing the two precue conditions for each participant revealed no significant difference, $t(7) = 0.185$ ($p = .858$).

Is there a part-report advantage for arrays of digits when selection is on the basis of temporal position?

The data in Figure 2a show the average performance of 12 participants, for each of the three conditions: part report no-delay, part report 500-ms cue delay, and whole report. For the part-report condition, the x-axis represents the delay from the offset of the second mask to the auditory cue. For all conditions, the y-axis represents the percentage (left-hand scale) and estimated number (right-hand scale) of digits available to the participant. For the whole-report conditions the estimated number of digits available is the sum of correctly reported digits across both arrays. In part-report conditions it is the number of digits reported per cued array, multiplied by 2 (given that there were two arrays, cued at random). Correct responses were recorded only when participants reported the correct digits in the locations in which they had occurred in the target arrays. The mean scores (shown in Figure 2a) were 5.12 digits (64.0%), 4.67 digits (58.4%), and 4.42 digits (55.3%), for the part report no delay, part-report 500-ms cue delay, and whole report, respectively. A one-way repeated measures analysis of variance (ANOVA) revealed a main effect of report condition, $F(2, 22) = 9.178$, $MSE = 0.023$, $p = .001$ (sphericity assumed). Since the overall ANOVA was significant, and there were only three groups, pairwise comparisons were made via Fisher's least significant difference (LSD; Howell, 1992, pp. 356) and revealed the following: a significant difference between the scores of the part-report conditions ($p = .015$); a significant difference between the part-report no-delay condition and the whole-report condition ($p = .002$); and no significant difference between the part-report delayed-cue condition and the whole-report condition ($p = .150$).

What is the relative distribution of digits reported from Array 1 and Array 2?

The main analysis considered performance irrespective of which of the two target arrays was cued. We now consider performance for the two arrays separately.

Average report performances per array are presented in Figure 2b. When there was no cue delay, performance was 63.7% and 64.3%, respectively, for trials on which the first or second array was cued, indicating that participants reported the two arrays equally well. With a 500-ms cue delay, performance was 52.7% and 64.1%, respectively, for trials on which the first or second array was cued. A two-way repeated measures ANOVA revealed a nonsignificant main effect of cued array, $F(1, 11) = 2.574$, $MSE = 0.043$, $p = .137$; a significant main effect of delay, $F(1, 11) = 8.193$, $MSE = 0.038$, $p = .015$; and a

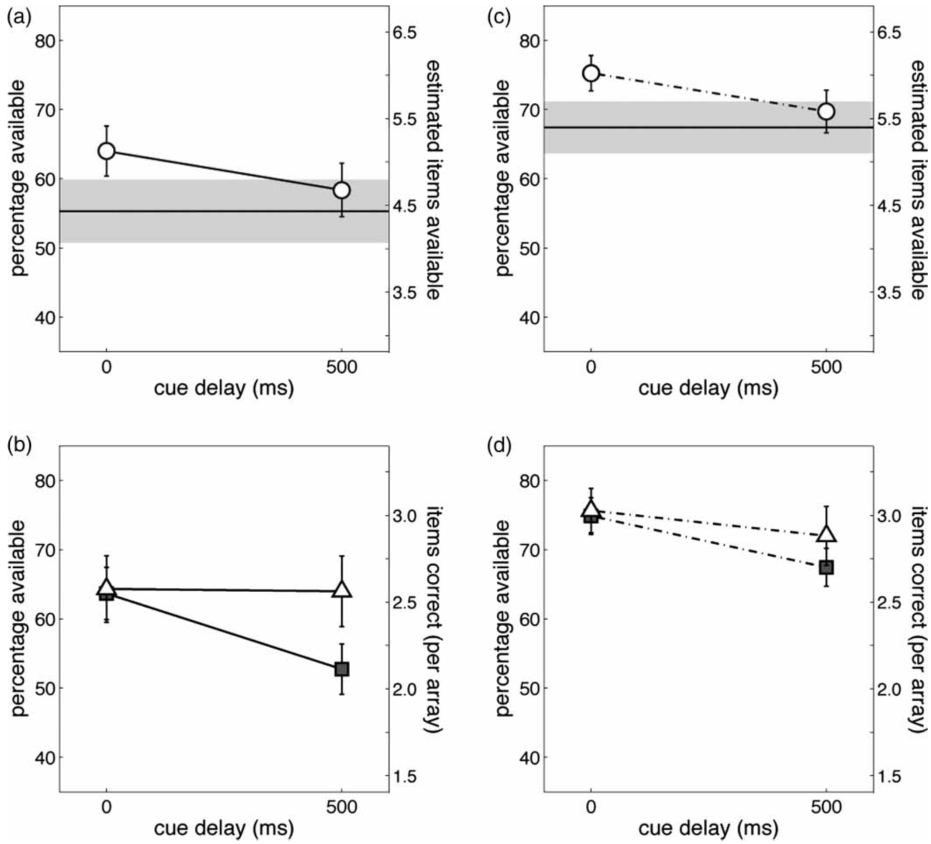

Figure 2. *(a) Average data collected from 12 participants from Experiment 1. The symbols represent data collected from the part-report conditions in which the delay between the second (final) mask and the auditory cue was either 0 or 500 ms. The horizontal line represents performance for the whole-report condition. In the whole-report condition the participants started reporting immediately after the presentation of the second mask. The y-axis on the left-hand side represents the percentage of correctly reported digits, and that on the right-hand side represents the estimated number of digits available to the participant. For the whole-report conditions, the estimated number of digits available is the sum of correctly reported digits across both arrays. For part-report conditions it is the number of digits reported per trial, multiplied by 2 (given that there were 2 possible cued arrays). The error bars associated with each symbol and the greyed region associated with the horizontal line are (±1 SE) across participants. (b) Average part-report performances analysed separately for each target array. The grey squares and open triangles represent the report performance for the first and second arrays, respectively. The y-axis on the left-hand side represents the percentage of correctly reported digits, and that on the right-hand side represents the number of digits available out of 4 for each array. Error bars are (±1 SE) across participants. (c) Average data when position information is ignored in scoring responses. The format is analogous to that of (a). (d) Average part-report performances for each target array when position information is ignored in scoring responses. The format is analogous to that of (b).*

nonsignificant interaction, $F(1, 11) = 3.932$, $MSE = 0.034$, $p = .073$. The analysis of simple effects showed a nonsignificant effect of delay for trials on which the second array was cued, $t(11) = 0.112$, $p = .913$, but a significant effect of delay for trials on which the first array was cued, $t(11) = 2.672$, $p = .022$.

Does the decline in performance with increasing cue delay reflect a loss of identity information or only a loss of spatial information?

Early work on iconic memory (Mewhort, Campbell, Marchetti, & Campbell, 1981; Townsend, 1973), particularly with masked displays (Mewhort, Marchetti, Gurnsey, &

Campbell, 1984), suggests that an increase in cue delay results in a loss of spatial information, while identity information remains preserved. We can address this issue by repeating our analysis but scoring a response correct if that item occurred anywhere within the cued array. So, for example, a response of 1234 to the target 4222 would receive a score of 2. In this system, once a response item has been paired with a stimulus item, both items are removed from further scoring. When location is important, chance performance is 1/10 (i.e., 0.4 letters out of 4); when location is not important, chance performance increases to 1.21 letters out of 4. To maintain consistent scoring between part- and whole-report, we counted as correct only transpositions within one spatial array, in both report types.

The mean scores calculated in this way (shown in Figure 2c) were 6.02 digits (75.3%), 5.58 digits (69.8%), and 5.39 digits (67.4%), for the part-report no-delay, part-report 500-ms cue delay, and whole report, respectively. A one-way repeated measures ANOVA revealed a main effect of report condition, $F(2, 22) = 10.35$, $MSE = 0.02$, $p = .001$ (sphericity assumed). Pairwise comparisons made via Fisher's least significant difference (LSD) revealed the following: a significant difference between the scores of the part-report conditions ($p = .004$); a significant difference between the part-report no-delay condition and the whole-report condition ($p = .002$); and no significant difference between the part-report delayed-cue condition and the whole-report condition ($p = .219$).

Average report performances per array are shown in Figure 2d. When there was no cue delay, performance was 74.8% and 75.7%, respectively, for trials on which the first or second array was cued, indicating that participants reported the two arrays equally well. With a 500-ms cue delay, performance was 67.5% and 72.0%, respectively, for trials on which the first or second array was cued. A two-way repeated measures ANOVA revealed a nonsignificant main effect of cued array, $F(1, 11) = 0.877$, $MSE = 0.009$, $p = .369$; a significant main effect of delay, $F(1, 11) = 13.219$, $MSE = 0.037$, $p = .004$; and a nonsignificant interaction, $F(1, 11) = 1.384$, $MSE = 0.004$, $p = .264$. The analysis of simple effects showed a nonsignificant effect of delay for trials on which the second array was cued, $t(11) = 1.623$, $p = .133$, but a significant effect of delay for trials on which the first array was cued, $t(11) = 3.409$, $p = .006$.

When positional information is ignored in scoring participants' responses, performance estimates are increased, consistent with the improvement expected from guessing. However, the overall pattern of results remains. In particular, the relatively poor performance for the first array with increased cue delay was still observed, suggesting that this impairment is unlikely to be due to a particular loss of spatial information.

Discussion

Experiment 1 measured the extent to which successively presented, spatially overlapping information can be stored and recalled. A whole-report condition was compared to a part-report condition in which an auditory postcue indicated which subset of information to report according to its temporal position in the stimulus sequence. Part-report performance with a cue presented immediately after the target sequence was higher than whole-report performance. The size of the advantage is 16%, a similar value to the part-report advantage reported by Darwin et al. (1972) in their classical study of echoic memory for verbal sequences, but a much smaller value than found by Sperling (1960) for spatial cueing of single visual arrays.

To explain the part-report advantage in our experiment, it might be suggested that the digits from the first array are immediately transferred to short-term memory before the auditory cue is presented and that the part-report advantage comes only from those trials in which the second array was cued. If this strategy were used by participants, it would be expected that report performance for information from the first array over both part-report conditions would be constant, and that performance when the first array was cued would not show a part-report superiority.

Our results are the opposite of this prediction: When the cue came immediately after the second mask, participants recalled items with equal accuracy from the first and second target arrays. As the cue delay was increased by a further 500 ms, the number of items recalled from the first array (relative to the second) declined, suggesting a decay of the information available from the first array over this period. An analysis that ignored the spatial position of items revealed the same pattern of results, suggesting that it is not only positional information that is lost when the cue is delayed.

The part-report advantage is compatible with the traditional definition of iconic memory and with current accounts of fragile VSTM, both of which describe a short-lived store with a larger capacity than short-term memory. The fact that this advantage is found when selection is on the basis of temporal position suggests that very-short-term visual memory can hold temporal-order information, and that sequentially presented information does not necessarily overwrite previously presented visual information. These properties are not normally incorporated in descriptions of iconic memory or of fragile VSTM, both of which suppose that only one temporal slice or snapshot of successive visual inputs is preserved. The following experiments are designed to test possible explanations of the results.

EXPERIMENT 2: EYE-MOVEMENT PATTERNS FOR THE PART-REPORT TASKS

During voluntary viewing of an object, the eyes are not stationary and continually move about the region of fixation. Even a highly trained participant exhibits microsaccades, drifts, and tremors during fixation, and some or all of these eye movements may have functional roles in vision (Carpenter, 1988; Martinez-Conde, Macknik, & Hubel, 2004).

Is it possible that in Experiment 1, either microsaccades or larger eye movements provide a way of recoding spatially overlapping sequentially presented visual information as spatially displaced retinal images? In Experiment 2, eye-movement patterns were measured during a replication of the part-report paradigm used in Experiment 1. We measured eye position during presentation of the first array and during presentation of the second array, and we used the difference between these measurements as an estimate of the displacement in eye position. A systematic shift in eye position that is correlated with task performance would suggest that participants can store overlapping temporally presented information in a traditional "snapshot" representation by introducing retinal displacement between successive temporal presentations

Method

Participants

There were 4 participants: 2 naïve (C.P. and A.F.) plus 2 authors W.S.S. and H.E.S.

Apparatus

Stimuli were presented on a 19-inch (0.475-m) ViewSonic monitor (displaying 1,024 × 768 pixels at a frame rate of 100 Hz). The graphics output was controlled via a CRS ViSaGe unit and a 4 × PCI express graphics card. Eye movements were recorded using a CRS 250-Hz infrared video-based eye tracker with an angular resolution of 0.05 deg. The display was gamma corrected (linearized) from measurements made with a CRS OptiCAL. The experiment was executed and controlled using the Matlab programming language (Version 6.5).

Stimuli

As with Experiment 1, target stimuli were 1-by-4 arrays of digits rendered in the "DS-Digital font". The height of each letter was fixed at 14 pixels (or 5 mm) subtending 0.64 degrees of visual angle at the viewing distance of 0.45 m. Each of the 4 digits in a target array was centred within the elements of an invisible grid such that the centre spacing between digits was fixed at 0.95 degrees. Checks within the mask stimuli were sized 3 × 3 pixels, the width of a digit-segment in the arrays.

The mask subtended 10.34 × 2.98 degrees of visual angle.

With the frame-rate of 100 Hz, and a frame-to-frame interval of 10 ms, the stimulus sequence was: Target Array 1 (20)−blank field (110)−mask (20)−blank field (110)−Target Array 2 (20)−blank field (110)−mask (20)−blank, where the numbers in parentheses represent the nominal duration of each item in milliseconds. Auditory cues were the same as those described for Experiment 1.

Procedure

Each participant obtained data for part-report no-delay and part-report delayed-cue conditions, with 32 trials per block. The order of conditions was randomized.

The procedure for Experiment 2 is the same as that described previously for the part-report conditions of Experiment 1, with the exception that eye position was monitored during the task and a calibration procedure was performed for the eye-tracking equipment using standard CRS calibration routine before the start of each experimental block.

Results

The average performances from 4 participants, shown in Figure 3a, were 5.69 digits (71.1%) and 5.05 digits (63.1%), for the part-report no-delay and part-report 500-ms delay conditions, respectively. Average report performances per array are presented in Figure 3b. When there was no cue delay, performance was 69.9% and 72.7%, respectively, for trials on which the first or second array was cued. With a 500-ms cue delay, performance was 56.3% and 69.5%, respectively, for trials on which the first or second array was cued.

A two-way repeated measures ANOVA on average scores per participant revealed a nonsignificant main effect of cued array, $F(1, 3) = 0.801$, $MSE = 0.026$, $p = .437$; a marginally significant main effect of delay, $F(1, 3) = 8.843$, $MSE = 0.028$, $p = .059$; and a significant interaction, $F(1, 3) = 62.253$, $MSE = 0.011$, $p = .004$. The analysis of simple effects showed a nonsignificant effect of delay for trials on which the second array was cued, $t(3) = 0.943$, $p = .415$, but a significant effect of delay for trials on which the first array was cued, $t(3) = 5.652$, $p = .011$.

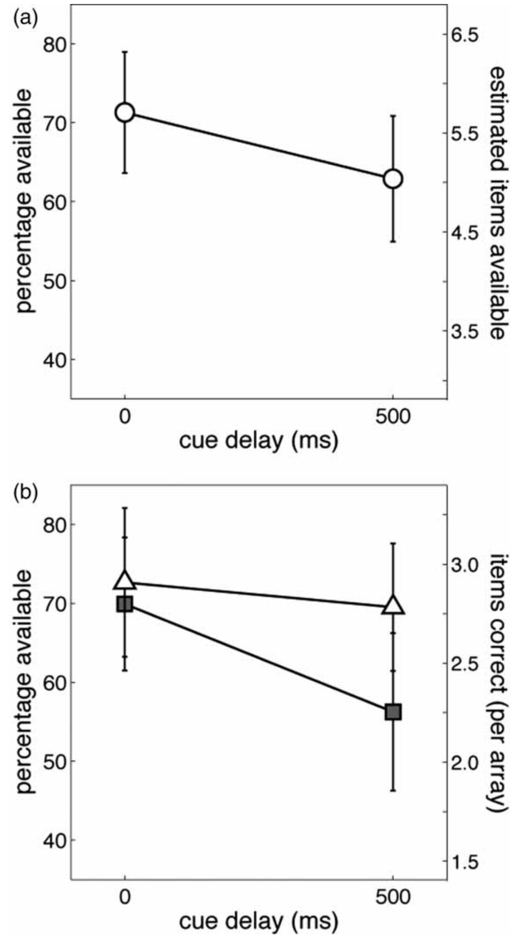

Figure 3. *(a) Average data collected from 4 participants from Experiment 2 in which participants' eye movements were recorded during the stimulus presentation. The format is the same as that for Figure 2 (a). (b) Average report performances for each target array. The format is the same as that for Figure 2 (b).*

Do participants shift their gaze between the presentation of arrays, and how does this relate to participant performance for the main task?

In this experiment, eye position was recorded at 250 Hz (1 sample every 4 ms). During the task, each digit array was present for two consecutive refreshes of the monitor (running at the frame rate of 100 Hz), so at least five samples of eye position were obtained for each digit array (assuming no loss of signal during recording). The average horizontal (x) and vertical (y) fixation positions (specified in pixels on the display) were derived from these multiple samples, providing a single (x, y) eye position per digit array. The Euclidian distance between eye positions for the two digit arrays was calculated, and this estimate of the magnitude of displacement between successive displays was converted to degrees of visual angle and used in subsequent analyses (see Figure 4).

Median magnitudes of displacement for the 4 participants were 0.64 (C.P.), 0.28 (W.S.S.), 0.33 (H.E.S.), and 0.29 (A.F.) degrees of visual angle. While displacements of this order of magnitude are similar to the spacing between digits (0.95 degrees of visual angle), they do not place the retinal image of the target array beyond the boundaries of the chequerboard masks that follow each target array.

In the next analysis we performed a more direct test of whether any displacement might have a functional benefit to task performance. On a trial-by-trial basis, we determined the correlation between task performance and the magnitude of retinal displacement between target arrays. We undertook a multiple regression, partitioning out variability between participants, as described in detail by Bland and Altman (1995). For both conditions (no-delay and delay) we obtained non-significant correlations, with $r = -.0593$, $F(1, 117) = 0.406$, $p = .525$ for no-delay, and $r = -.0861$, $F(1, 110) = 0.818$, $p = .368$ for delay.

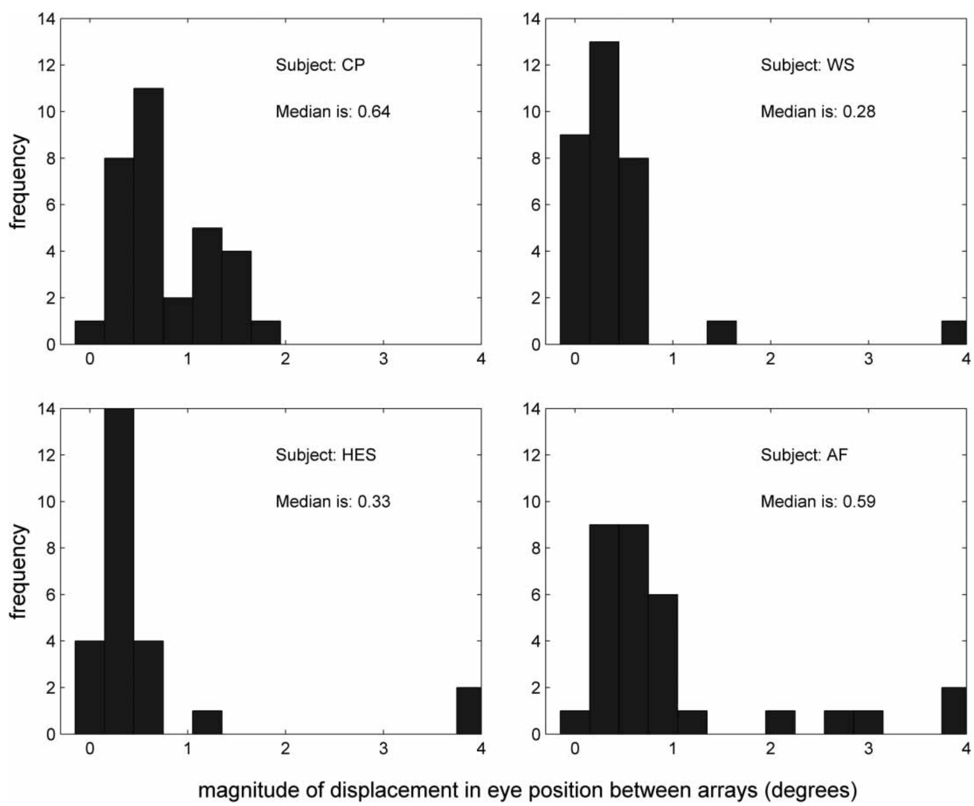

Figure 4. *Eye-movement data collected from 4 participants for the part-report task with no cue delay. Each graph shows the distribution of the magnitudes of eye position displacements measured between the first and second digit arrays for individual participants.*

In each case, there was no interaction between participants and the covariate ($p = .701$ and $p = .885$, respectively), implying that the derived correlations are representative of all participants. These results indicate that there is not a significant relationship between the participant's score on a trial and eye-movement displacement between target arrays, suggesting that a shift in eye gaze between presentation of the first and second arrays does not either improve or impair recall of the digits.

Discussion

In Experiment 2, eye movements were recorded whilst part-report performance of temporally presented visual information was measured. As with Experiment 1, the results support the idea that very-short-term visual memory can retain successively presented information.

Measurements of eye movements during the task confirm that a decaying partial-report superiority is found when a sequence of stimuli fall on spatially overlapping regions of the retina: Median displacements in eye position were not sufficient to separate the spatially extended stimuli.

Multiple linear regression analysis revealed that differences in eye position between presentations of target arrays are not correlated with participant report performance. This suggests that shifting gaze between presentations of the two digit arrays (thereby placing them on a different retinal area) is not instrumental in improving recall of information from the cued target array.

EXPERIMENT 3: INTERLEAVED PRESENTATION OF DIFFERENT REPORT CONDITIONS

In Experiment 1, the three different report conditions were presented in different blocks. It could be argued that participants would adopt different strategies in different blocks. For example, in the whole report condition, the optimum strategy is to begin transferring items to a more durable store immediately after the first array is presented, whereas in the part report conditions the transfer should be delayed until the cue is presented. Potential differences in strategic set become particularly important when we consider performance differences for the two arrays. For example, might the interaction between cue delay and array number in Figures 2b and 3b reflect not a change over time but a change in strategy?

In Experiment 3, the train of visual stimuli was the same as in Experiment 1 but we interleaved the three report conditions within each block of trials.

Method

Participants

There were 4 participants: R.L. (practised but naïve to the nature of the experiment), T.P. (inexperienced and naïve to the nature of the experiment), plus 2 authors, W.S.S. and H.E.S.

Apparatus and stimuli

All apparatus and stimulus properties were identical to those described for Experiment 1 with the exception of the auditory cues. Three auditory tones indicated which task the participants were required to complete: a high tone (1.7 kHz) indicated part-report from the first array; a low tone (441 Hz) indicated part-report from the second array; and an intermediate tone (882 Hz) indicated whole report. The visual stimulus sequence was preceded by a sequence of three warning tones of the same pitch as the tone of intermediate frequency.

Procedure

Each participant obtained data in three sessions, each containing 96 trials, 32 for each of the three report conditions: whole report, part-report no-delay, and part-report delayed cue. The order of conditions was randomized, so that each type of postcue had a probability of one third on every trial. All other details of the procedure were the same as those for Experiment 1.

Results

The mean scores, averaged across repetitions and across observers (shown in Figure 5a) were 5.48 digits (68.5%), 4.94 digits (61.8%), and 3.99 digits (49.9%), for the part-report no-delay, part-report 500-ms delay, and whole report, respectively, indicating a part-report advantage of 37% between part-report no-delay and whole report. We performed a 3×3 repeated measures ANOVA, with repetition number and report condition as factors. The ANOVA (sphericity assumed) revealed a main effect of report conditions, $F(2, 6) = 53.858$, $MSE = 0.106$, $p < .001$, no main effect of repetition, $F(2, 6) = 3.929$, $MSE = 0.006$, $p = .081$, and no interaction, $F(4, 12) = 2.559$, $MSE = 0.002$, $p = .093$. Pairwise comparisons between report conditions were made via Fisher's least significant difference (LSD) and revealed the following: a significant difference between the part-report conditions ($p = .009$), a significant difference between the part-report no-delay condition and the whole-report condition ($p = .005$), and a significant difference between the part-report delayed-cue condition and the whole-report condition ($p = .005$).

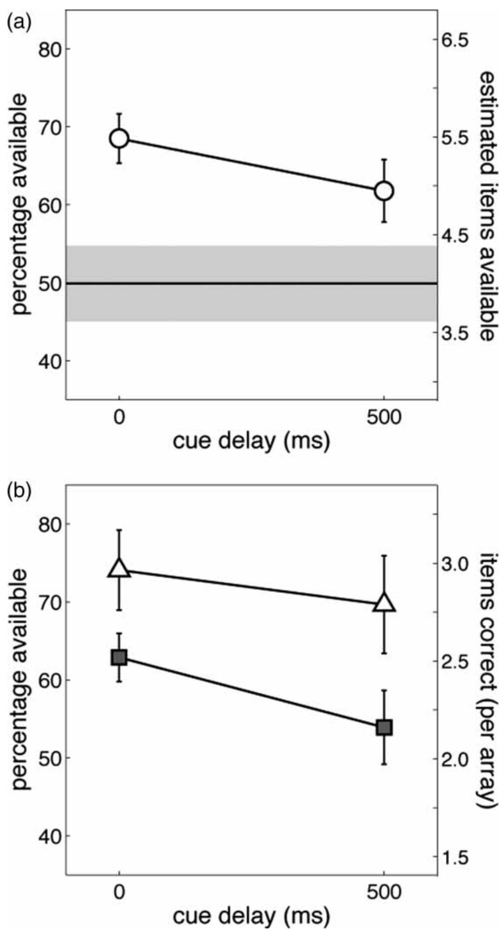

Figure 5. *(a). Average data collected from 4 participants from Experiment 3 in which trials from the three experimental conditions (whole report, part-report no-delay, and part-report delayed-cue) were interleaved. The format is the same as that for Figure 2 (a). (b) Average report performances for each target array. The format is the same as that for Figure 2 (b).*

Average report performances per array are presented in Figure 5b. When there was no cue delay, performance was 62.9% and 74.1%, respectively, for trials on which the first or second array was cued. With a 500-ms cue delay, performance was 53.9% and 70.0%, respectively, for trials on which the first or second array was cued.

A two-way repeated measures ANOVA on scores per participant averaged over repetitions revealed a nonsignificant main effect of cued array, $F(1, 3) = 4.449$, $MSE = 0.073$, $p = .125$; a significant main effect of delay, $F(1, 3) = 35.424$, $MSE = 0.018$, $p = .009$; and a nonsignificant interaction, $F(1, 3) = 1.185$, $MSE = 0.002$, $p = .356$. The analysis of simple effects showed a nonsignificant effect of delay when the second array was cued, $t(3) = 1.595$, $p = .209$, but a significant effect of delay when the first array was cued, $t(3) = 4.738$, $p = .018$. As with Experiment 1, the performance improvement with an early cue reflects an improvement in participants' ability to report digits from the first array.

Discussion

In Experiment 3 we used an interleaved design in which, for each trial, any of the three report conditions (part-report no-delay, part-report delay, whole report) were equally probable. This design maintains the same strategic set for each report condition. The best performance was obtained for trials in which the part-report cue was presented immediately after the visual stimulus train, and when the cue occurred 500 ms after the stimulus train, performance declined. It is worth explicitly noting that delaying the cue by 500 ms causes a reduction in performance on the first array, even in these interleaved conditions when strategic set is maintained between trials with different cue delays. At the moment of the early cue, it seems that information from the first array remains available in a labile store, even though the array has been followed by a mask, another target, and a further mask.

The whole-report performance was worse than that of either part-report condition, indicating a loss of information during the reporting period. In blocked whole-report trials the reporting process could be triggered on presentation of the arrays. However, in the interleaved case, this strategy is unavailable, and so whole-report performance in Experiment 3 is likely to be lower than that in Experiment 1, as our results indicate.

In sum, the results of Experiment 1 suggest that information from both arrays is held for a short time in a visual memory, the contents of which can be selectively accessed according to an aftercoming cue that prompts report from either the first or the second array. Experiment 3 shows that the decline in performance with increasing cue delay is not dependent on a change in strategy between reporting conditions.

EXPERIMENT 4: REDUCING TARGET DURATION RELATIVE TO MASK DURATION

The results of the previous experiments imply that participants are able to store information about visually presented digits (from both target arrays) in a short-lived memory. The 7-segment font and binary noise masks were used to minimize the possibility that information could be recovered through a single combined visual image in which all stimuli in the sequence are integrated, but the digits still readable. If the part-report advantage in Experiments 1, 2, and 3 relied on recovering information from an integrated (short-lived, high-capacity) representation of the targets and masks, then decreasing the duration of the target array relative to the mask should decrease the part-report advantage. This is because a decrease in target duration relative to that of the mask means that an integrated representation is weighted more to the mask than the target; and this in turn would make the digit information more difficult to retrieve from an integrated representation. A reduction in part-report advantage with this timing configuration would suggest that participants may be using an integrated representation of the stimuli to do the task. If the pattern of report performance is unchanged then further support would be gained for the interpretation presented previously, that successive elements in the stimulus sequence are stored independently.

With this rationale, Experiment 1 was repeated with a reduction in the duration of the target arrays. To establish operationally that the shortened target was unreadable when integrated with the mask, we performed a control experiment in which a one-frame target array was followed immediately by a three-frame mask, with no interposed uniform field.

Experiment 3, and its control experiment, tested whether participants can still show a part-report advantage when it is not possible for any information to be recovered via an integrated representation of the digit-array and mask. If a part-report advantage is found under these conditions with relatively weak targets, it supports the suggestion made previously that successive inputs are not inescapably combined in their stored representation.

Method

Participants

For the main conditions and precue conditions there were 4 participants: R.L. (practised but naïve to the nature of the experiment), T.P. (inexperienced and naïve to the nature of the experiment), plus 2 authors, W.S.S. and H.E.S. The two most practised observers (authors W.S.S. and H.E.S.) participated in the control task.

Apparatus and stimuli

All apparatus and stimulus properties were identical to those described for Experiment 1, with the exception of the stimulus durations. For the main task, two target arrays comprising 4 digits were presented for 1 frame (at 140 Hz, giving a frame-to-frame interval of 7.14 ms), and each was followed by a uniform field for 15 frames and a binary noise mask for 3 frames. The sequence was: Target Array 1 (7)−blank field (107)−mask (21)−blank field (107)−Target Array 2 (7)−blank field (107)−mask (21)−blank, where the numbers in parentheses represent the nominal duration of each item in milliseconds. For the control task one target array comprising 4 digits was presented for 1 frame followed immediately by a binary noise mask for 3 frames. The sequence for this experiment was: target array (7)−mask (21), where the numbers in parentheses represent the nominal duration of each item in milliseconds.

Procedure

For the main task, each participant obtained data in blocked conditions for whole report, part-report no-delay, and part-report delayed cue, with 32 trials per block. The order of conditions was randomized, and the set was repeated three times. In addition, each participant completed one block of 32 precue trials, with 16 trials cued to the first array and 16 cued to the second array.

In addition, 2 participants completed 32 trials of the control task, in which a single array was presented (for 1 frame) and followed immediately by a mask (presented for 3 frames): Participants were asked to report as many as possible of the digits from the single array and to guess when they were unsure of the correct answer.

Results

For the control task, the average score for 32 trials for participant H.E.S. was 9% and for participant W.S.S. was 11%. The binomial probability of successful guessing is 10%, with standard deviation 3%.

Averages for the two precue conditions, in which stimulus timings were the same as those for the main task, and participants were instructed in advance to attend either to the first or to the second array, were 3.06 digits (out of a possible 4, 76.6%), and 2.91 digits (out of a possible 4, 72.7%) respectively. A paired t test comparing the two precue conditions for each participant revealed no significant differences, $t(3) = 0.476$, $p = .667$.

For the main task, the mean scores (shown in Figure 6a) were 4.52 digits (56.5%), 4.01 digits (50.1%), and 3.89 digits (48.6%), for the

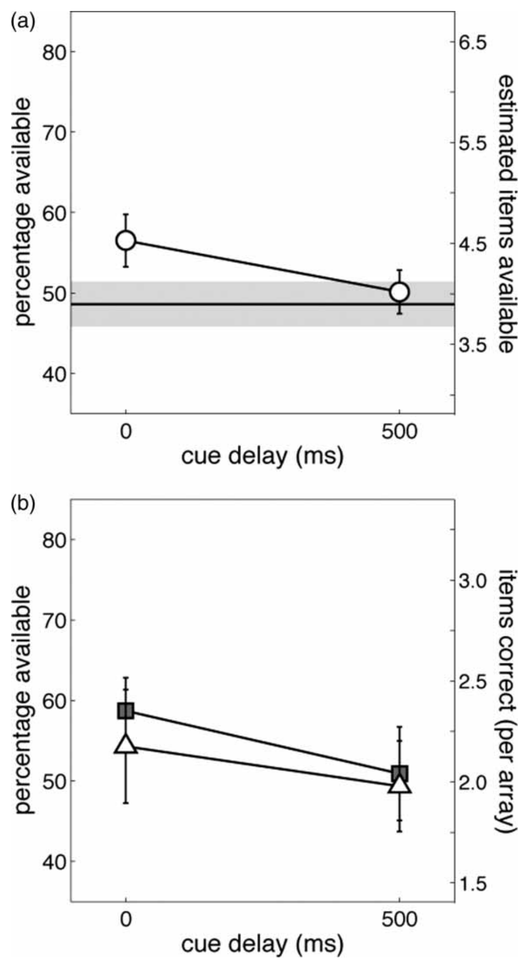

Figure 6. *(a) Average data collected from 4 participants from Experiment 4 in which the duration of the targets was reduced relative to the duration of the chequerboard masks. The format is the same as that for Figure 2 (a). (b) Average report performances for each target array. The format is the same as that for Figure 2 (b).*

part-report no-delay, part-report 500-ms delay, and whole report, respectively, indicating a part-report advantage of 16% between part-report no-delay and whole report. A two-way 3 × 3 repeated measures ANOVA (sphericity assumed) revealed a main effect of report conditions, $F(2, 6) = 9.721$, $MSE = 0.021$, $p = .013$, no main effect of repetition, $F(2, 6) = 2.452$, $MSE = 0.005$, $p = .167$, and no interaction, $F(4, 12) = 0.643$, $MSE = 0.001$, $p = .642$. Pairwise comparisons between report conditions were made via Fisher's least significant difference (LSD) and revealed the following: a significant difference between the part-report conditions ($p = .022$), a marginally significant difference between the part-report no-delay condition and the whole-report condition ($p = .055$), and no significant difference between the part-report delayed-cue condition and the whole-report condition ($p = .361$).

Average report-performances per array are presented in Figure 6b. When there was no cue delay, performance was 58.7% and 54.3%, respectively, for trials on which the first or second array was cued. With a 500-ms cue delay, performance was 50.9% and 49.4%, respectively, for trials on which the first or second array was cued.

A two-way repeated measures ANOVA on scores per participant averaged over repetitions revealed a nonsignificant main effect of cued array, $F(1, 3) = 0.095$, $MSE = 0.004$, $p = .778$; a significant main effect of delay, $F(1, 3) = 18.907$, $MSE = 0.016$, $p = .022$; and a nonsignificant interaction, $F(1, 3) = 1.288$, $MSE = 0.001$, $p = .339$. The analysis of simple effects showed a nonsignificant effect of delay for trials on which the second array was cued, $t(3) = 2.420$, $p = .094$, but a significant effect of delay for trials on which the first array was cued, $t(3) = 4.302$, $p = .023$.

Discussion

In Experiment 4, the targets were presented only for a single frame and were followed (after a delay of 15 frames) by masks that had a duration of 3 frames, providing a much higher effective contrast. The results are consistent with those of Experiment 1. Both experiments show a part-report advantage for temporally cued digits when the postcue is presented without delay. Experiment 4 confirms that this part-report advantage survives when the target arrays are presented for only a very brief duration and are therefore relatively weak.

The performance values of all conditions of Experiment 4 are lower than those for Experiment 1 but this may be due to a decrease in the readability of the digits at such a brief duration, an explanation that is consistent with the reduction in performance for the two precue conditions, which measure the extent to which the digits from each array are readable.

For the control condition, the 1-frame target and 3-frame mask were presented with no delay between them, so that the total sequence duration was 4 frames of the display monitor (~30 ms). When target and mask are presented in this configuration they perceptually combine, and participants are unable to read any digits: Performance fell to chance level of 10%.

Taken together, the experiments presented in this section suggest that the part-report advantage for information from the cued array is not derived from a combined representation of the mask and target arrays.

EXPERIMENT 5: VOCAL REHEARSAL DURING TASK

The preceding experiments suggest that participants can store and retrieve sequentially presented information. A postcue instructing participants to select information on the basis of temporal position was sufficient to support a part-report advantage provided the cue was presented sufficiently early.

Many studies have demonstrated recoding of visually presented linguistic material into an auditory form during memory tasks (Baddeley, 1966, 2000). Is it possible that in our experiments successive items are not held within the perceptual buffer in sequence, but that a subset of digits are held in a phonological store while others are recovered from a conventional icon? For example, the participant might recode the first array into verbal auditory representations that can be maintained by the articulatory loop, but might read out items directly from the icon (or fragile VSTM) if the second array is cued. In the experiments presented so far, there are two arguments against this latter suggestion. First, the second array is itself followed by a chequerboard mask and would therefore not be available for retrieval from a conventional icon. Secondly, the analysis of all preceding experiments showed a more rapid decay of material from the first array than from the second array, a result that would not be expected if the first array were held in a phonological store and maintained by rehearsal. Here we use articulatory suppression to occupy the articulatory loop and so to test explicitly whether the part-report advantage we measure arises because the participant has available to him a phonological store in which he can place a subset of the material (Baddeley, Thomson, & Buchanan, 1975).

Method

Participants

There were 4 participants, R.L. (practised but naïve to the nature of the experiment), T.P. (inexperienced and naïve to the nature of the experiment), and authors W.S.S. and H.E.S.

Apparatus and stimuli

All apparatus and stimuli were identical to those described for Experiment 1.

Procedure

The procedure for Experiment 5 was the same as that described for Experiments 1, 3, and 4 with the exception that participants were required to rehearse verbally the digits 0–9, in order, during presentation of the stimuli. Each participant obtained data for whole-report, part-report no-delay, and part-report delayed-cue conditions, with 32 trials per block. The order of conditions was randomized, and the set was repeated three times. In addition, each participant completed one block of 32 precue trials, with 16 trials cued to the first array and 16 cued to the second array.

The cue to participants to start rehearsing was the fixation cross in the centre of the display screen, which appeared prior to the onset of the first target array (see description of Experiment 1). Participants were required to stop rehearsing when they heard the auditory cue after the final mask stimulus. In the case of the part-report condition, this auditory cue indicated which array to report. In the case of the whole report, the auditory cue indicated when to stop rehearsal and to start reporting all eight target digits. Participants typically completed three to four cycles of rehearsal, and the spoken digits were clearly enunciated and recognizable to the experimenter.

Results

Averages for the two precue conditions in which participants were required to attend to and report only the first or second array, were 87.1% (i.e., 3.48 digits out of a possible 4) and 88.3% (i.e., 3.53 digits out of a possible 4), respectively. A paired t test comparing the two precue conditions for each participant revealed no significant differences, $t(3) = 0.397$ ($p = .718$).

The mean scores (shown in Figure 7a) were 5.53 digits (69.1%), 4.89 digits (61.1%), and 4.65 digits (58.1%), for the part report no-delay, part-report 500-ms delay, and whole report, respectively, indicating a part-report advantage of 19% between part-report no-delay and whole report. A two-way 3 × 3 repeated measures ANOVA (sphericity assumed) revealed a main effect of report conditions, $F(2, 6) = 7.651$, $MSE = 0.039$, $p = .022$, and a main effect of repetition, $F(2, 6) = 11.441$, $p = .009$, but did not reveal an interaction, $F(4, 12) = 0.917$, $MSE = 0.002$, $p = .485$.

Pairwise comparisons between report conditions were made via Fisher's least significant difference (LSD) and revealed the following: a significant difference between the part-report conditions ($p = .022$), a nonsignificant difference between the part-report no-delay and the whole-

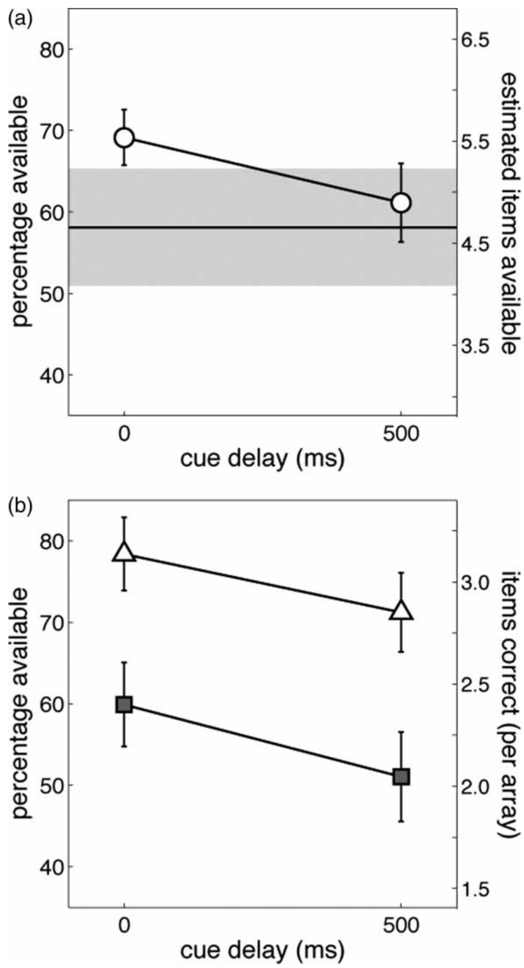

Figure 7. *(a) Average data collected from 4 participants from Experiment 5 in which participants were engaged in vocal rehearsal during presentation of the stimulus train. The format is the same as that for Figure 2 (a). (b) Average report performances for each target array. The format is the same as that for Figure 2 (b).*

report conditions ($p = .072$), and no significant difference between the part-report delayed-cue condition and the whole-report condition ($p = .296$).

Average report performances per array are presented in Figure 7b. When there was no cue delay, performance was 59.9% and 78.4%, respectively, for trials on which the first or second array was cued. With a 500-ms cue delay, performance was 51.0% and 71.2%, respectively, for trials on which the first or second array was cued.

A three-way repeated measures ANOVA (Cued Array × Cue Delay × Repetition) revealed a significant main effect of cued array, $F(1, 3) = 13.671$, $MSE = 0.449$, $p = .034$; a significant main effect of delay, $F(1, 3) = 19.143$, $MSE = 0.077$, $p = .022$; and a nonsignificant main effect of repetition, $F(2, 6) = 4.078$, $MSE = 0.012$, $p = .076$. None of the two- or three-way interactions were significant. The analysis of simple effects showed a marginally significant effect of delay for trials on which the second array was cued, $t(3) = 3.072$, $p = .054$, and a significant effect of delay for trials on which the first array was cued, $t(3) = 3.241$, $p = .048$.

Discussion

The results of Experiment 5 again demonstrate decline in overall part-report performance when the cue is delayed by 500 ms. If the performance advantage with an early cue were achieved by transfer of a subset of the material to a phonological store, we should have expected the effect to be abolished by coarticulation. In this experiment, the difference between part-report no-delay and whole-report conditions failed to reach significance. We asked participants to vocally rehearse digits whilst presenting digits visually for verbal report. Using the same set of items for the primary task and for the suppression task may have complicated the results; performance in whole-report blocks was particularly variable in this experiment owing perhaps to patterns of interference and enhancement between the two tasks.

EXPERIMENT 6: INTERLEAVED PART- AND WHOLE-REPORT CONDITIONS WITH WEAK TARGETS AND VOCAL REHEARSAL

How secure is the part-report advantage for temporal cueing to successive visual inputs? In the final experiment we combined several of the previous manipulations. We used brief targets, to reduce their relative strength and thus prevent digits being recovered from a single integrated representation of targets and masks. We used concurrent articulation to occupy the phonological loop and ensure that performance depended on recovery of items from visual storage. And we randomly interleaved three types of trial (whole-report, and part-report with and without cue delay) to prevent participants adopting different strategies in different conditions.

Method

Participants

There were 9 participants: the authors R.B. and H.E.S., plus 7 naïve participants, two of whom were practised on the task.

Apparatus and stimuli

The apparatus and stimuli were identical to those for the main task of Experiment 4 with the following exceptions. The auditory cues were modified to improve their rapid interpretation. A single click (2.81 kHz) indicated part-report from the first array; a double click (2.81 kHz, separated by 38 ms) indicated part-report from the second array; and a low tone (441 Hz) indicated whole report. The duration of the clicks was 2.5 ms and that of the tone was 20 ms. The sequence of visual stimuli was preceded by two warning tones, which had the same pitch as the low-frequency tone and were separated by 1 s.

Procedure

The procedure for Experiment 6 was the same as that described for the main task of Experiment 3 with the exception that participants were required to rehearse aloud "*bla-bla*", during presentation of the stimuli (Larsen & Baddeley, 2003). They typically completed 7 to 9 clearly enunciated repeats of "*bla-bla*". A second change was that participants now entered their responses via a wireless number-pad, instead of reporting them verbally. Each participant obtained data in five sessions, each containing 150 trials (30 each for whole report and for the four combinations of target array and delay during part report). The order of trials within a session was randomized. The first session was considered as practice, and the data were not included in the analyses.

Results

A total of two participants were able to report only 1 digit correctly on part-report trials. Therefore, they were excluded after completing two runs as they did not show improvement. These participants did complain that they found the first array particularly difficult to see, even when instructed to attend only to the first array. All other participants were able to do the task. On a few trials participants entered fewer digits than required. This happened only for 0.18 % of the trials when the participants had to report 4 digits, and 7.86 % of the trials when they had to report all the 8 digits.

The mean scores, (shown in Figure 8a) were 4.39 digits (54.9%), 4.02 digits (50.3%) and 3.38 digits (42.3%), for part-report no-delay, part-report 500-ms delay, and whole report, respectively, giving a part-report advantage of 29.9% between part-report no-delay and whole report. A two-way 3 × 3 repeated measures ANOVA (sphericity assumed) revealed a main effect of report conditions, $F(2, 12) = 16.872$, $MSE = 7.373$, $p < .0001$, and an effect of repetition, $F(3, 18) = 3.641$, $MSE = 0.759$, $p = .03$, but no significant interaction, $F(6, 36) = 1.401$, $MSE = 0.130$, $p = .241$.

Pairwise comparisons between report conditions made via Fisher's least significant difference (LSD) showed significant differences between the part-report conditions ($p = .013$), between the part-report no-delay condition and the whole-report condition ($p = .003$), and between the part-report delayed-cue condition and the whole-report condition ($p = .017$).

Average report performances per array are presented in Figure 8b. When there was no cue delay, performance was 59.3% and 50.5%, respectively, for trials on which the first or second array was cued. With a 500-ms cue delay, the corresponding values were 52.8% and 47.7%.

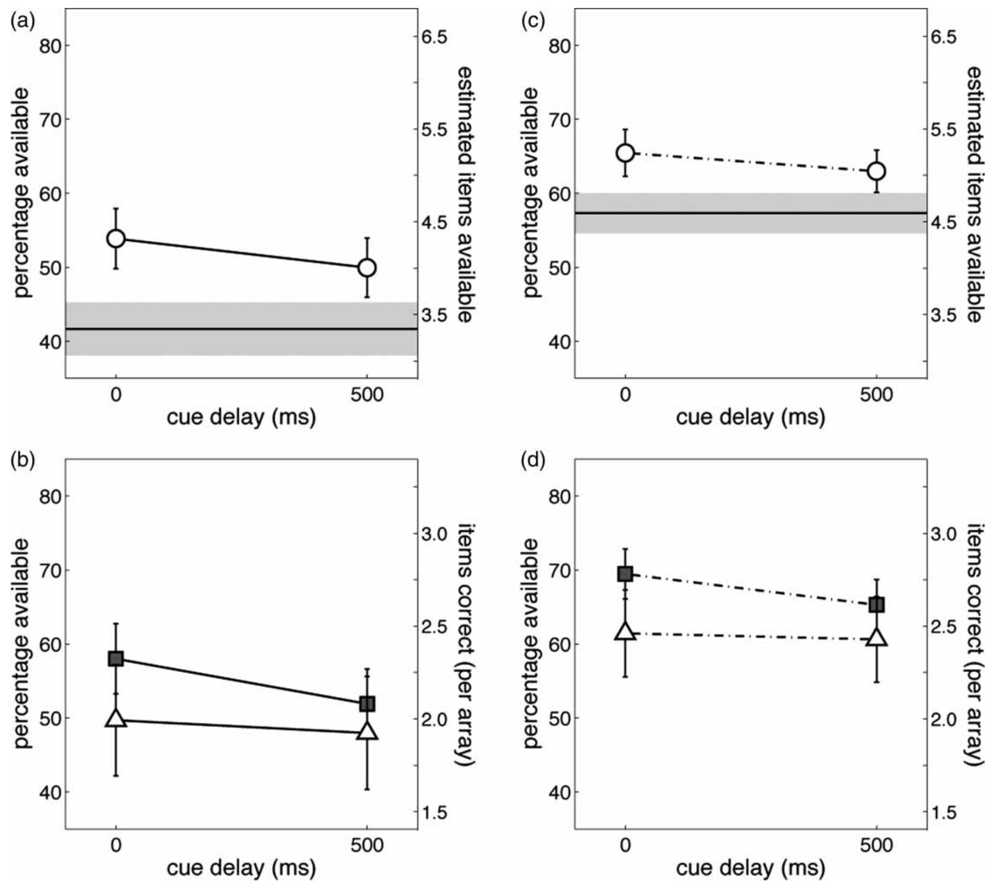

Figure 8. *(a) Average data collected from 7 participants from Experiment 6 in which trials from the three experimental conditions were interleaved, the duration of the targets was reduced relative to the duration of the chequerboard masks, and participants were engaged in vocal rehearsal during presentation of the stimulus train. The format is the same as that for Figure 2 (a). (b) Average report performances for each target array. The format is the same as that for Figure 2 (b). (c) Average data when position information is ignored in scoring responses. The format is analogous to that of (a). (d) Average part-report performances for each target array when position information is ignored in scoring responses. The format is analogous to that of (b).*

A three-way repeated measures ANOVA (Cued Array × Cue Delay × Repetition) revealed no significant main effect of cued array, $F(1, 6) = 0.519$, $MSE = 2.167$, $p = .498$; a significant effect of delay, $F(1, 6) = 13.317$, $MSE = 0.966$, $p = .013$; no significant effect of repetition, $F(3, 18) = 2.203$, $MSE = 0.239$, $p = .123$; and a marginally significant repetition by delay interaction, $F(3, 18) = 3.085$, $MSE = 0.096$, $p = .054$. None of the other two- or three-way interactions were significant. The analysis of simple effects showed no significant effect of delay when the second array was cued, $t(6) = 1.253$, $p = .257$, but a significant effect of delay when the first array was cued, $t(6) = 3.082$, $p = .022$.

To evaluate if the part-report advantage with and without cue delay was due to a loss of spatial information, rather than identity information, we repeated our analysis by scoring a response correct if that item occurred anywhere within the cued array (as was done for Experiment 1). The mean scores, averaged across repetitions and across participants, (shown in Figure 8c) were now 5.31 digits (66.4%), 5.05 digits (63.1%), and

4.62 digits (57.8%), for part-report no-delay, part-report 500-ms delay, and whole report, respectively. A two-way 3 × 3 repeated measures ANOVA (sphericity assumed) revealed a significant effect of report conditions, $F(2, 12) = 21.837$, $MSE = 3.394$, $p < .0001$, no significant effect of repetition, $F(3, 18) = 2.396$, $MSE = 0.359$, $p = .102$, and no significant interaction, $F(6, 36) = 0.969$, $MSE = 0.059$, $p = .460$. Pairwise comparisons between report conditions made via Fisher's least significant difference (LSD) revealed significant differences between the two part-report conditions ($p = .010$), between the part-report no-delay condition and the whole-report condition ($p = .002$), and between the part-report delayed-cue condition and the whole-report condition ($p = .008$).

Average performances when spatial information is ignored are shown for each array in Figure 8d. When there was no cue delay, performance was 70.4% and 62.2%, respectively, for trials on which the first or second array was cued. With a 500-ms cue delay, the corresponding values were 65.5% and 60.8%.

A two-way repeated measures ANOVA on scores per participant averaged over repetitions revealed no significant main effect of cued array, $F(1, 6) = 0.795$, $MSE = 0.465$, $p = .407$; a significant effect of delay, $F(1, 6) = 13.543$, $MSE = 0.112$, $p = .010$; and a nonsignificant interaction, $F(1, 6) = 2.515$, $MSE = 0.034$, $p = .164$. The analysis of simple effects showed a nonsignificant effect of delay when the second array was cued, $t(6) = 0.939$, $p = .384$, but a significant effect of delay when the first array was cued, $t(6) = 3.775$, $p = .01$.

Discussion

Experiment 6 again demonstrates a part-report advantage for temporally cued arrays when the postcue is presented without delay. The advantage declines significantly when the cue is delayed, but remains significantly above whole-report performance. As in all the experiments, delaying the cue impairs performance for items recovered from the first array. Experiment 6 shows that blocking the phonological loop does not eliminate this short-lived part-report advantage for visual information that has been followed by a mask, a second target array, and another mask.

GENERAL DISCUSSION

We set out to test whether there exists a visual buffer that can preserve successive items in their temporal sequence. The experiments consistently showed a part-report advantage when an after-coming cue instructed the participant to select on the basis of temporal position in the train of stimuli. Delaying the cue by 500 ms significantly reduces the performance level, moving it towards that in whole-report conditions. This rapid decline in part-report superiority has been taken as one of the signatures of sensory storage.

We place most weight on the following argument: In every one of the experiments, performance on the first array is better if it is cued immediately after the second array and the second mask, rather than after a further delay of 500 ms. This implies that, although two masks and the second array have intervened, a fragile representation of the first array is still available when the earlier cue is interpreted. A further 500 ms later, little can be recovered from this fragile representation. The decline in performance on the first array is not what one would expect if the observer were transferring the contents of the first array to a phonological store, while recovering the second array from a conventional icon. The results from Experiment 6, in which concurrent articulatory suppression was used, further supports the argument that observers were not using a phonological store during our tasks.

The present experiments extend the finding of Smithson and Mollon (2006) who showed that observers could select from an array of letters on the basis of a spatial cue, even when that cue followed the presentation of a noise mask. In that former experiment, the successive stimuli were separated in time and differed in low-level spatial characteristics, and the selection was on the basis of spatial position. In the present experiment, the two target arrays share the same statistical properties, and retrospective selection must be made purely on the basis of time of presentation. In the earlier experiment, passive filtering on the basis of low-level features might have prevented integration of the representations of target and mask, but in the present experiments the two arrays are identical in their low-level features, and so the first array could not be preserved by such a mechanism. Moreover, since the two arrays are identical in spatial position and in their visual features, we infer that there is a fragile representation of both target arrays within which temporal order is preserved.

Iconic storage versus fragile VSTM

The high capacity iconic store has traditionally been distinguished from the more stable, lower capacity VSTM. Operationally these two stores differ in their durations—iconic memory is held to survive for only a few hundred milliseconds, whereas VSTM may survive for several seconds (Phillips, 1974). The stores also differ in their capacity, in that iconic memory can hold in excess of a dozen items (Sperling, 1960), whereas VSTM is limited to approximately four items (Luck & Vogel, 1997). A further difference is that iconic storage is held to occur passively, whereas information in VSTM can be actively maintained (Kroll, Parks, Parkinson, Bieber, & Johnson, 1970; Phillips, 1974).

However, it is less straightforward to distinguish between iconic memory and the fragile form of VSTM: Both putative stores exhibit part-report superiority effects, and both are vulnerable to an after-coming pattern mask (Sligte et al., 2008). It is also difficult to distinguish fragile VSTM from the early form of visual representation postulated by Rensink (2000): In his three-stage scheme, volatile proto-objects are formed rapidly and in parallel across the visual scene, and focused attention acts to crystallize a limited subset of these in a stable form, until the withdrawal of attention causes the stable forms to dissolve once more to proto-objects.

Few of the existing accounts of early visual storage allow for storage of successive items: Iconic memory is held to be overwritten by after-coming pattern masks (Phillips, 1974; Sperling, 1960); fragile VSTM is disrupted by new patterned inputs (Sligte et al., 2008); and Rensink's proto-objects are "*replaced* when any new stimulus appears at their retinal location" (Rensink, 2000, p. 20). In postulating a short-lived visual buffer that can hold successive inputs, we do not wish to add a further store to those mentioned above. Rather we suggest that existing models might be modified to accommodate our results.

Properties of the postulated store

We postulate a rapidly decaying store capable of preserving visual events in sequence in the same way as successive phonemes are preserved in an auditory echoic store. Treisman et al. (1975) have previously shown that the icon can hold spatio-temporal information in the form of motion, but this result could be accounted for by dedicated motion detectors relatively early in the visual system; our results suggest a stronger sense in which successive items are preserved independently. Attention can be directed to a particular position in the temporal sequence (Schill & Zetzsche, 1995), and this can be done on the basis of an after-coming cue.

One recurrent feature of our results is that delaying the cue causes a rather limited decline in performance with the second array. Any traditional model that assumes an exponentially decaying passive icon would predict that the loss over a fixed time period would be more marked closer in time to the stimulus presentation. In our experiments the cue is closest in time to the more recent target, so we might expect performance on the second target to decline more rapidly. However, if the visual sequence is preserved in a fragile visual memory, our result rather makes sense, since attention may adhere most strongly to the last target event.

Echoic memory, like early visual memory, is thought to have a large capacity, the purpose being to buffer incoming sensory information before limited attentional resources can be directed to it at a later time. The partial-report superiority that we measure is of a similar magnitude to that demonstrated with auditory sequences and suggests that a buffered sequence of visual inputs can be accessed for a brief time. We have not formally tested how many discrete arrays may be held, but we know that there are limitations imposed by the duration of the storage and the resolution at which discrete items can be represented within the store. It is also possible that the spatio-temporal buffer evolved not for holding discrete alphanumeric arrays, but for holding stimuli that vary continuously in time and that exhibit smooth spatial transitions (Allik & Bachmiann, 1983; Johansson, 1983; Klatzky, 1983; Mollon, 1969; Phillips, 1983). Examples of such stimuli would be gestures, gaits, and trajectories, and it is possible that the postulated store might exhibit a larger capacity for inputs that vary continuously in space and time than for discrete items.

Stimuli in our visual environment vary in four dimensions: three spatial dimensions and one temporal dimension. Are all four dimensions represented in the postulated buffer? In the experiments reported here, we provide an explicit test of whether temporal position is represented in the buffer, but do not test for simultaneous representation of three spatial dimensions. Xu and Nakayama (2007) have shown that performance is enhanced in a spatial VSTM task if the memoranda are distributed between two depth planes. It is possible that the additional storage capacity available for depth can be diverted to represent time in our task. Our participants reported the subjective impression that the successive arrays often seemed to occupy different depth positions in their internal representation. A combination of cueing in time and in three-dimensional space could probe the extent to which spatial and temporal dimensions are represented independently.

The mechanism of temporal storage

Our experiments provide behavioural evidence that a temporal-order cue can be used to report selectively from a short-lived store that concurrently preserves information about successive inputs. It is possible to envisage two types of mechanism that could underlie such storage.

There is growing evidence that the neural circuits responsible for precise sensory encoding are also responsible for the maintenance of internal representations of sensory stimuli (Pasternak & Greenlee, 2005). So one proposal for the maintenance of successive inputs would be that the apparatus of feature analysis is reduplicated and that successive inputs are assigned to the different sets of analysers, in a cycling fashion. The number of reduplicated sets of analysers would set the capacity of the temporal store. To account for the estimated temporal capacity of the echoic store (approximately 2 seconds; Baddeley et al., 1975) several duplicated batteries of analysers would be needed in the auditory case. The reduplication might be sidestepped if, as Barlow (2007) has argued from considerations of efficiency, an individual neuron has the computational power to briefly store the spatio-temporal pattern of its inputs, as opposed to instantaneous snapshots.

Alternatively, objects might be represented as a more abstract feature list, as postulated, for example, by Wheeler and Treisman (2002). Kahneman, Treisman, and Gibbs (1992) suggested that such object descriptions carried with them location tags. The representation of time is inherently no more complex than the representation of space. Perhaps analogous tags are used in brief visual storage to encode the temporal location of successive objects.

Original manuscript received 4 December 2008
Accepted revision received 13 July 2010
First published online 6 January 2011

REFERENCES

Allik, J., & Bachmiann, T. (1983). How bad is the icon? *Behavioral and Brain Sciences*, *6*(1), 12–13.

Averbach, E., & Coriell, A. S. (1961). Short-term memory in vision. *The Bell System Technical Journal*, *40*(1), 309–328.

Baddeley, A. D. (1966). The influence of acoustic and semantic similarity on long-term memory for word sequences. *Quarterly Journal of Experimental Psychology*, *18*(4), 302–309.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *14*(6), 575–589.

Barlow, H. (2007). The nested networks of brains and minds. In G. R. Bock & J. A. Goode (Eds.), *The limits of reductionism in biology* (pp. 142–159). Chichester, UK: John Wiley.

Becker, M. W., Pashler, H., & Anstis, S. M. (2000). The role of iconic memory in change-detection tasks. *Perception*, *29*(3), 273–286.

Bland, J. M., & Altman, D. G. (1995). Statistics notes: Calculating correlation coefficients with repeated observations: Part 1. Correlation within subjects. *British Medical Journal*, *310*(6977), 446.

Bridgeman, B. (1998). Durations of stimuli displayed on video display terminals: $(n - 1)/f +$ persistence. *Psychological Science*, *9*, 232–233.

Carpenter, R.H. S. (Ed.). (1988). *Movements of the eyes* (2nd ed.). London: Pion.

Chow, S. L. (1986). Iconic memory, location information, and partial report. *Journal of Experimental Psychology: Human Perception and Performance*, *12*(4), 455–465.

Coltheart, M. (1980). Iconic memory and visible persistence. *Perception and Psychophysics*, *27*(3), 183–228.

Coltheart, M. (1983). Iconic memory. *Philosophical Transactions of the Royal Society, B*, *302*(1110), 283–294.

Darwin, C. J., Turvey, M. T., & Crowder, R. G. (1972). An auditory analogue of the Sperling partial report procedure: Evidence for a brief auditory storage. *Cognitive Psychology*, *3*(2), 255–267.

Freyd, J. J. (1987). Dynamic mental representations. *Psychological Review*, *94*(4), 427–438.

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

Griffin, I. C., & Nobre, A. C. (2003). Orienting attention to locations in internal representations. *Journal of Cognitive Neuroscience*, *15*(8), 1176–1194.

Guttman, N., & Julesz, B. (1963). Lower limits of auditory periodicity analysis. *Journal of the Acoustical Society of America*, *35*(4), 610.

Howell, D. C. (1992). *Statistical methods for psychology*. Belmont, CA: Duxbury Press.

Jiang, Y., & Kumar, A. (2004). Visual short-term memory for two sequential arrays. One integrated representation or two separate representations? *Psychonomic Bulletin and Review*, *11*(3), 495–500.

Johansson, G. (1983). Optic flow, icons, and memory. *Behavioral and Brain Sciences*, *6*(1), 23–24.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, *24*. 175−219.

Keysers, C., Xiao, D.-K. Földiák, P., & Perrett, D. I. (2005). Out of sight but not out of mind: The neurophysiology of iconic memory in the superior temporal sulcus. *Cognitive Neuropsychology*, *22*(3−4), 316–332.

Klatzky, R. L. (1983). The icon is dead: Long live the icon. *Behavioral and Brain Sciences*, *6*(1), 27–28.

Kroll, N. E., Parks, T., Parkinson, S. R., Bieber, S. L., & Johnson, A. L. (1970). Short-term memory while shadowing: Recall of visually and of aurally presented letters. *Journal of Experimental Psychology*, *85*(2), 220–224.

Landman, R., Spekreijse, H., & Lamme, V. A. F. (2003). Large capacity storage of integrated objects before change blindness. *Vision Research*, *43*(2), 149–164.

Larsen, J. D., & Baddeley, A. (2003). Disruption of verbal STM by irrelevant speech, articulatory suppression, and manual tapping: Do they have a common source? *Quarterly Journal of Experimental Psychology*, *56*(8), 1249–1268.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281.

Makovski, T., & Jiang, Y. V. (2007). Distributing versus focusing attention in visual short-term memory. *Psychonomic Bulletin and Review*, *14*(6), 1072–1078.

Martinez-Conde, S., Macknik, S. L., & Hubel, D. H. (2004). The role of fixational eye movements in visual perception. *Nature Reviews Neuroscience*, *5*(3), 229–240.

Mewhort, D. J. K., Campbell, A. J., Marchetti, F. M., & Campbell, J. I. D. (1981). Identification, localization, and iconic memory—an evaluation of the bar-probe task. *Memory and Cognition*, *9*(1), 50–67.

Mewhort, D. J. K., Marchetti, F. M., Gurnsey, R., & Campbell, A. J. (1984). Information persistence: A dual-buffer model for initial processing. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 287–298). Hove, UK: Lawrence Erlbaum Associates.

Mollon, J. D. (1969). *Temporal factors in perception*. Unpublished DPhil dissertation, University of Oxford, Oxford, UK.

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

Palmer, J. (1988). Very short-term visual memory for size and shape. *Perception and Psychophysics*, *43*(3), 278–286.

Pasternak, T., & Greenlee, M. W. (2005). Working memory in primate sensory systems. *Nature Reviews Neuroscience*, *6*(2), 97–107.

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception and Psychophysics*, *16*(2), 283–290.

Phillips, W. A. (1983). Change perception needs sensory storage. *Behavioral and Brain Sciences*, *6*(1), 35–36.

Rensink, R. A. (2000). The dynamic representation of scenes. *Visual Cognition*, *7*(1), 17–42.

Robbins, S. E. (2004). On time, memory and dynamic form. *Consciousness and Cognition*, *13*(4), 762–788.

Schill, K., & Zetzsche, C. (1995). A model of visual spatio-temporal memory: The icon revisited. *Psychological Research*, *57*(2), 88–102.

Sligte, I. G., Scholte, H. S., & Lamme, V. A. F. (2008). Are there multiple visual short-term memory stores? *PLoS ONE*, *3*(2), e1699.

Smithson, H. E., & Mollon, J. D. (2006). Do masks terminate the icon? *Quarterly Journal of Experimental Psychology*, *59*(1), 150–160.

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

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

Stockman, A., Langendörfer, M., Smithson, H. E., & Sharpe, L. T. (2006). Human cone light adaptation: From behavioral measurements to molecular mechanisms. *Journal of Vision*, *6*, 1194–1213.

Tatler, B. W. (2001). Characterising the visual buffer: Real-world evidence for overwriting early in each fixation. *Perception*, *30*(8), 993–1006.

Townsend, V. M. (1973). Loss of spatial and identity information following a tachistoscopic exposure. *Journal of Experimental Psychology*, *98*(1), 113–118.

Treisman, A., Russell, R., & Green, J. (1975). Brief visual storage of shape and movement. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and Performance V* (pp. 699–721). London: Academic Press.

Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, *131*(1), 48–64.

Xu, Y. D., & Nakayama, K. (2007). Visual short-term memory benefit for objects on different 3-D surfaces. *Journal of Experimental Psychology: General*, *136*(4), 653–662.