Vision Research 96 (2014) 106–112

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Is visual short-term memory depthful?

Adam Reeves [*], Quan Lei (雷泉)

Department of Psychology, 125 NI, Northeastern University, Boston, MA 02115, USA

ARTICLE INFO

Article history:
Received 21 August 2013
Received in revised form 15 January 2014
Available online 1 February 2014

Keywords:
Visual short-term memory
Iconic memory
Partial report
Depth

ABSTRACT

Does visual short-term memory (VSTM) depend on depth, as it might be if information was stored in more than one depth layer? Depth is critical in natural viewing and might be expected to affect retention, but whether this is so is currently unknown. Cued partial reports of letter arrays (Sperling, 1960) were measured up to 700 ms after display termination. Adding stereoscopic depth hardly affected VSTM capacity or decay inferred from total errors. The pattern of transposition errors (letters reported from an uncued row) was almost independent of depth and cue delay. We conclude that VSTM is effectively two-dimensional.

1. Introduction

Although a great deal has been learned about visual short-term memory (VSTM) storage since Sperling (1960), the visual displays that have been used to test VSTM, such as letters arrayed in rows and columns, have been presented in the picture plane and lack any variation in depth. Indeed, the locution 'icon' (Neisser, 1967) for VSTM and the terminology 'iconic decay' suggest that most authors have implicitly assumed that the mental representation is indeed 2-D, given that an 'icon' is a flat painting of a religious figure or subject seen from straight on with little or no depth modeling. Clearly, if the input image is flat, as with the 2-D letter array of Sperling (1960), it is hardly surprising that the representation of it should also exclude variation in depth. However, in natural viewing, different objects normally occupy different depths as well as different spatial positions. The difference between normal viewing and picture-plane viewing was discussed extensively by Gibson (1979); suffice it to say here that natural viewing provides more sources of information to the perceiver than picture-plane viewing. We therefore wondered if adding depth to the traditional flat letter-array stimuli might affect VSTM. An advantage of letter-array stimuli is that they make contact with the extensive iconic memory literature, results of which have been taken to define the properties of VSTM. The resulting 3-D displays have rows, columns, and discrete planes, however, and do not vary continuously in space or time. Thus they only inch towards the naturalistic percepts discussed by Gibson (1979), and this limitation should be kept in mind.

Given the importance of depth in visual perception, one might expect it to affect VSTM. For example, VSTM might be layered, such that the typical information limit found with a single depth plane could be by-passed or reduced if information were distributed across multiple depth planes. Indeed, recall of foreground information might proceed relatively independently of recall of middle-distance or background information. We created depth by using stereoscopic disparity since visual information can be perceived in multiple depth planes when using disparity, as shown by Julesz (1971). Disparity is processed more rapidly than the icon decays, the integration time for stereopsis being about 100 ms with vergence controlled (Harwerth, Fredenburg, & Smith, 2003). Thus having multiple disparate planes in the image could *in principal* affect the contents of VSTM. In the only relevant study we could find, Xu and Nakayama (2007) discovered a small improvement in recall after a 2 s delay for visual information portrayed on more than one disparity-generated surface compared to information portrayed on a single surface. We wondered if this effect might be a consequence of a larger, more meaningful difference in decay rates at earlier times. Decay might be slower for information in multiple depth planes than for information in a single depth, if depth supported information in VSTM, even if the asymptotic capacities were similar. In contrast, VSTM might automatically store 'depth tags' indicating the distance of an object to the perceiver. If so, the capacity of VSTM might be reduced by the addition of such tags, so even if depth affects VSTM, improvement is not the only possibility.

Since partial-report experiments necessarily involve both transfer of information to VSTM and shifts of attention, we looked for a model which might help us interpret any depth effects we discovered. As Reeves and Sperling (1986) had found time constants for attention shifts between 133 and 183 ms, somewhat faster than the quarter-second decay of the icon, variations in attention shift latency might well affect partial report. Gegenfurtner and Sperling

---

[*] Corresponding author. Fax: +1 6173738714.
E-mail address: reeves@neu.edu (A. Reeves).

http://dx.doi.org/10.1016/j.visres.2014.01.010
0042-6989/

(1993) showed that prior to a cue, subjects primarily attend to the center of the letter array, but after the cue, they attend only to the cued row, a 'center-out' strategy. In their model, transfer rate to VSTM was the product of iconic legibility (which depends on time) and attentional allocation (which shifts after a cue). This model has implications for our experiments. We chose display conditions such that depth would have no effect on iconic legibility, the letters being as identical in the 'flat' as in the 'depth' condition, and being widely separated in space to avoid lateral masking or crowding. Thus individual letters should decay at the same rate whether presented in a 2-D or 3-D context. Therefore any effect of depth must either be on attentional allocation or on storage of the array in VSTM. Our plan was therefore to ascertain if there is a depth effect, and if there was, to determine which mechanism was responsible.

A potential complication arises in testing the role of depth because shifting attention within a depth plane may be faster than shifting attention between depth planes. Downing and Pinker (1985) showed that reaction time was slower for targets that were at a different depth plane from a cued location. Atchley et al. (1997) found that attention in 3-D space functions like a spotlight which is extended in depth as well as in the horizontal and the vertical dimensions. Andersen and Kramer (1993) asked subjects to report a target letter (X or O) flanked by compatible or incompatible distractors. The increase in response latency due to distractor incompatibility was greatest when the distractors were portrayed stereoscopically on the same depth plane as the target, and dropped off as disparity was increased to 6 minarc. The drop-off was fastest when the distractors were portrayed behind the target rather than in front, perhaps because far-to-near attention shifts are faster than near-to-far shifts (Arnott & Shedden, 2000; Downing & Pinker, 1985.) If shifting attention across depth planes is slower than shifting within a single picture plane, then performance with multiple depth planes may be worse than with a single depth plane, since attention must be shifted to the cued row in the partial report paradigm and the icon will have decayed further while attention was shifting. Thus, finding performance differences between 2-D and 3-D displays may not imply that depth is encoded in VSTM, only that attention is culpable. However, Iavecchia and Folk (1994), who used a spatial cuing task, found no difference between the time course of within-plane and across-plane attention shifts. Ghirardelli and Folk (1996) found no cost for switching attention in depth, when the target appeared at a cued or uncued depth. Since of the studies mentioned, two showed no effect and the others only small effects, we anticipate that this complication would not obscure a major role for depth in VSTM.

An alternative prediction, of no depth effect, stems from the several studies by Sakitt and colleagues (Long & Sakitt, 1980; Sakitt & Long, 1979; Sakitt, 1975, 1976). They accounted for iconic memory in terms of retinal function. Their specific claim that rods determine iconic persistence and cones determine perceived offset (Sakitt & Long, 1978) was falsified using cone-only presentation conditions (Adelson, 1978), but any similar retinal basis would imply that stereoscopic depth in the letter display could not affect VSTM storage, setting aside the attention-shift complication just discussed. Depth might still affect report, but only at a stage of processing subsequent to storage, such as selection of items for retrieval. A similar logic would apply if, for example, VSTM depends on a flat representation in visual cortex (Nikolić et al., 2009).

2. General method

2.1. Participants

All participants were undergraduates enrolled in an introductory psychology course at Northeastern University. They gave informed consent and participated in the experiments for course credit. All had normal stereopsis as screened with a Julesz random-dot stereogram, and normal (20/20) or corrected-to-normal vision in both eyes. One potential participant without stereopsis, and twelve with poorer acuity in either eye, were excluded from the experiment. The procedures were approved by the Northeastern University IRB.

2.2. Apparatus

Stimuli were generated by custom MATLAB routines and presented on a 19 in. diagonal CRT monitor viewed from 57 cm. The display has a resolution of 1024 × 768 pixels, a refresh rate of 100 Hz, and was driven by a Cambridge Research Systems VSG-5 card programmed in Matlab V.6 under Windows XP. The VSG card provides accurate timing of display frames when run repeatedly in 'movie' mode, as confirmed with a counter triggered by a photodiode: every 10 ms frame was timed correctly over a 20 min. calibration period. Stimulus chromaticity was (0.290, 0.300) in CIE (x, y) co-ordinates as recorded with a calibrated Cambridge Research Systems colorimeter. Stimulus luminance was 116 cd/m^2. Stimuli appeared white on a black background. A Wheatstone stereoscope arrangement was used to produce stereoscopic stimuli. The screen was divided in half by a cardboard speculum extending from the nose to the screen. A 20-diopter wedge-shaped prism was placed in a holder in front of the left eye so that the left- and right-hand images could be superimposed easily while the participant verged on the surface of the CRT monitor. Every participant who passed the Julesz RDS and acuity screens reported experiencing depth with this arrangement.

2.3. Procedure

Stimuli were upper-case letters randomly selected from the alphabet and organized in rows of 3 letters each. Letters on the same row always had the same disparity and same size (Fig. 1). The partial-report technique was employed to test participants' memory for letters under different display conditions. Each trial started with a fixation display consisting of crosses occupying

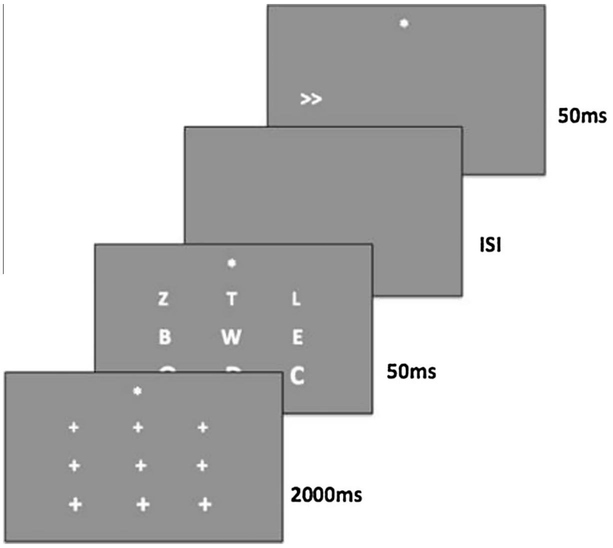

Fig. 1. A post-cue trial in Experiment 1. An initial fixation array, shown for 2 s, is replaced by the letter array for 50 ms, and then, after a blank ISI, by an arrow cue for 50 ms. The letter and cue arrays are in reverse order in pre-cue trials. Smaller letters are at top in both flat and depth conditions. In depth, the larger letters at bottom were brought forward by adding stereoscopic disparity.