

# Visual Short Term Memory

Authored by  
**mohammad looti**

June 14, 2026

## RECOMMENDED CITATION

mohammad looti (2026). *Visual Short Term Memory*. PSYCHOLOGICAL SCALES.  
Retrieved from <https://scales.arabpsychology.com/?p=38254>

In the study of vision, visual short-term memory (VSTM) is one of three broad memory systems including iconic memory and long-term memory. VSTM is a type of short-term memory, but one limited to information within the visual domain.

The term VSTM refers in a theory-neutral manner to the non-permanent storage of visual information over an extended period of time. The Visuospatial Sketchpad is a VSTM subcomponent within the theoretical model of working memory proposed by Alan Baddeley. Whereas iconic memories are fragile, decay rapidly, and are unable to be actively maintained, visual short-term memories are robust to subsequent stimuli and last over many seconds. VSTM is distinguished from long-term memory, on the other hand, primarily by its very limited capacity.

## Overview

The introduction of stimuli which were hard to verbalize, and unlikely to be held in long-term memory, revolutionized the study of VSTM in the early 1970s (Cermak, 1971; Phillips, 1974; Phillips & Baddeley, 1971). The basic experimental technique used required observers to indicate whether two matrices (Phillips, 1974; Phillips & Baddeley, 1971), or figures (Cermak, 1971), separated by a short temporal interval, were the same. The finding that observers were able to report that a change had occurred, at levels significantly above chance, indicated that they were able to encode some aspect of the first stimulus in a purely visual store, at least for the period until the presentation of the second stimulus. However, as the stimuli used were complex, and the nature of the change relatively uncontrolled, these experiments left open various questions, such as: (1) whether only a subset of the perceptual dimensions comprising a visual stimulus are stored (e.g., spatial frequency, luminance, or contrast); (2) whether some perceptual dimensions are maintained in VSTM with greater fidelity than others; and (3) the nature by which these dimensions are encoded (i.e., are perceptual dimensions encoded within separate, parallel channels, or are all perceptual dimensions stored as a single bound entity within VSTM?).

## Set-size effects in VSTM

Much effort has been dedicated to investigating the capacity limits of VSTM. In a typical change-detection task, observers are presented with two arrays, composed of a number of stimuli. The two arrays are separated by a short temporal interval, and the task of observers is to decide if the first and second arrays are identical, or whether one item differs across the two displays (e.g., Luck & Vogel, 1997). Performance is critically dependent on the number of items in the array. While performance is generally almost perfect for arrays of one or two items, correct responses invariably decline in a monotonic fashion as more items are added. Different theoretical models have been put forward to explain limits on VSTM storage, and distinguishing between them remains an active area of research.

## Models of capacity limits in VSTM

### Slot models

A prominent class of model proposes that observers are limited by the total number of items which can be encoded, either because the capacity of VSTM itself is limited (e.g., Cowan, 2001; Luck & Vogel, 1997; Pashler, 1988). This type of model has obvious similarities to urn models used in probability theory (see, for example, Mendenhall, 1967). In essence, an urn model assumes that VSTM is restricted in storage capacity to only a few items,  $k$  (often estimated to lie in the range of three-to-five in adults, though fewer in children (Riggs, McTaggart & Simpson, 2006)). The probability that a suprathreshold change will be detected is simply the probability that the change element is encoded in VSTM (i.e.,  $k/N$ ). This capacity limit has been linked to the posterior parietal cortex, the activity of which initially increases with the number of stimuli in the arrays, but saturates at higher set-sizes (Todd & Marois, 2004). Although urn models are used commonly to describe performance limitations in VSTM (e.g., Luck & Vogel, 1997; Pashler, 1988; Sperling, 1960), it is only recently that the actual structure of items stored has been considered. Luck and colleagues have reported a series of experiments designed specifically to elucidate the structure of information held in VSTM (Luck & Vogel, 1997). This work provides evidence that items stored in VSTM are coherent objects, and not the more elementary features of which those objects are composed.

### Noise models

An alternative framework has more recently been put forward by Wilken and Ma (2004) who suggest that apparent capacity limitations in VSTM are caused by a monotonic decline in the quality of the internal representations stored (i.e., monotonic increase in noise) as a function of set-size. In this conception capacity limitations in memory are not caused by a limit on the number of things that can be encoded, but by a decline in the quality of the representation of each thing as more things are added to memory. In their 2004 experiments, they varied color, spatial frequency, and orientation of objects stored in VSTM using a signal detection theory approach (see also the closely related work by Palmer, 1990). The participants were asked to report differences between the visual stimuli presented to them in consecutive order. The investigators found that different stimuli were encoded independently and in parallel, and that the major factor limiting report performance was neuronal noise (which is a function of visual set-size).

Under this framework, the key limiting factor on working memory performance is the precision with which visual information can be stored, not the number of items that can be remembered. Further evidence for this theory was obtained by Bays and Husain (2008) using a discrimination task. They showed that, unlike a "slot" model of VSTM, a signal-detection model could account both for discrimination performance in their study and previous results from change detection tasks (e.g. Luck and Vogel, 1997). These authors proposed that VSTM is a flexible resource, shared out

between elements of a visual scene--items that receive more resource are stored with greater precision. In support of this, they showed that increasing the salience of one item in a memory array led to that item being recalled with increased resolution, but at the cost of reducing resolution of storage for the other items in the display.

### **Psychophysical models**

Psychophysical experiments suggest that information is encoded in VSTM across multiple parallel channels, each channel associated with a particular perceptual attribute (Magnussen, 2000). Within this framework, a decrease in an observer's ability to detect a change with increasing set-size can be attributed to two different processes: (1) if decisions are made across different channels, decreases in performance are typically small, and consistent with decreases expected when making multiple independent decisions (Greenlee & Thomas, 1993; Vincent & Regan, 1995); (2) if multiple decisions are made within the same channel, the decrease in performance is much greater than expected on the basis of increased decision-noise alone, and is attributed to interference caused by multiple decisions within the same perceptual channel (Magnussen & Greenlee, 1997).

However, the Greenlee-Thomas model (Greenlee & Thomas, 1993) suffers from two failings as a model for the effects of set-size in VSTM. First, it has only been empirically tested with displays composed of one or two elements. It has been shown repeatedly in various experimental paradigms that set-size effects differ for displays composed of a relatively small number of elements (i.e., 4 items or less), and those associated with larger displays (i.e., more than 4 items). The Greenlee-Thomas (1993) model offers no explanation for why this might be so. Second, while Magnussen, Greenlee, and Thomas (1997) are able to use this model to predict that greater interference will be found when dual decisions are made within the same perceptual dimension, rather than across different perceptual dimensions, this prediction lacks quantitative rigor, and is unable to accurately anticipate the size of the threshold increase, or give a detailed explanation of its underlying causes.

In addition to the Greenlee-Thomas model (Greenlee & Thomas, 1993), there are two other prominent approaches for describing set-size effects in VSTM. These two approaches can be referred to as sample size models (Palmer, 1990), and urn models (e.g., Pashler, 1988). They differ from the Greenlee-Thomas (1993) model by: (1) ascribing the root cause of set-size effects to a stage prior to decision making; and (2) making no theoretical distinction between decisions made in the same, or across different, perceptual dimensions.

### **Intermediate visual store**

A recent study has obtained evidence that could support the existence of an intermediate visual

store with characteristics of both iconic memory and VSTM (Sligte et al., 2008). This intermediate store is proposed to have high capacity (up to 15 items) and prolonged memory trace duration (up to 4 seconds). It coexists with VSTM but unlike it visual stimuli can overwrite the contents of its visual store. Further research suggests an involvement of visual area V4 (Sligte et al., 2009).

ARABPSYCHOLOGY.COM