Available online at www.sciencedirect.com
ScienceDirect
Cognitive Psychology 56 (2008) 237–283
Cognitive Psychology
www.elsevier.com/locate/cogpsych

Dynamic binding of identity and location information: A serial model of multiple identity tracking

Lauri Oksama [a,b,*], Jukka Hyönä [b]

[a] *Department of Behavioral Sciences, National Defence College, University of Turku, FIN-20014 Turku, Finland*
[b] *Department of Psychology, University of Turku, FIN-20014 Turku, Finland*

Accepted 6 March 2007
Available online 23 April 2007

Abstract

Tracking of multiple moving objects is commonly assumed to be carried out by a fixed-capacity parallel mechanism. The present study proposes a serial model (MOMIT) to explain performance accuracy in the maintenance of multiple moving objects with distinct identities. A serial refresh mechanism is postulated, which makes recourse to continuous attention switching, a capacity-limited episodic buffer for identity-location bindings, indexed location information stored in the visuospatial short-term memory, and an active role of long-term memory. As identity-location bindings are refreshed serially, a location error is inherent for all other targets except the focally attended one. The magnitude of this location error is a key factor in predicting tracking accuracy. MOMIT's predictions were supported by the data of five experiments: performance accuracy decreased as a function of target set-size, speed, and familiarity. A mathematical version of MOMIT fitted nicely to the observed data with plausible parameter estimates for the binding capacity and refresh time.

Keywords: Multiple identity tracking; Dynamic visual scenes; Identity-location binding; Episodic buffer; Visuospatial short-term memory; Long-term memory; Change detection; Multiple object tracking

* Corresponding author. Fax: +358 2 3335060.
E-mail addresses: loksama@utu.fi (L. Oksama), hyona@utu.fi (J. Hyönä).

0010-0285/$ - see front matter
doi:10.1016/j.cogpsych.2007.03.001

1. Introduction

Our perceptual-attentional system faces a challenging task when attempting to simultaneously track multiple moving elements in a visual scene. Tasks of this kind are common in many real-life visual environments, such as in traffic and sports, for example, when a player keeps track of other moving players during a soccer game, or when an air traffic controller monitors aircraft on a radar screen. In order to make quick and sensible decisions in tasks like these (e.g., passing the ball to a team mate, or giving ATC clearance to aircraft), observers need to keep track of where each visual element is located at any given time.

Computationally, in this kind of task observers must be able to dynamically bind correct identities to continuously changing spatiotemporal locations: observers face a dynamic what–where binding problem (for the binding problem, see e.g., Treisman, 1996; the special issue of Visual Cognition 3–5/8, 2001). In the real world, our perceptual–attentional system must solve this problem in very different environments and stimulus conditions. The target stimuli may vary in many ways with respect to, for example, their spatiotemporal and semantic properties. The number of objects-to-be-tracked may vary and so may speed, movement direction, and movement trajectory of the target objects, as well as the number of distracter items present. The objects may temporarily disappear due to external (occlusion) or internal (saccadic suppression) reasons. The objects may be familiar or unfamiliar to the observer. Finally, the tracked objects may be visually identical to or distinct from each other.

There is empirical evidence demonstrating that binding of 'what' to 'where' in a visual scene is not an easy task for the attentional-perceptual system even when the observed scenes are static. For example, data suggest that observers may be prone to illusory conjunctions, that is, to erroneously bind visual feature combinations (e.g., binding the red color with a triangle when it should have been attached to a circle; Treisman & Schmidt, 1982). Moreover, it seems that what–where bindings are not achieved automatically in an early stage of visual processing. Instead, neurophysiological evidence suggests that the visual input is functionally decomposed into separate dimensions in the ventral (what) and the dorsal (where) streams (e.g., Ungerleider & Mishkin, 1982). All this supports the idea that the dynamic binding problem described above is real and that it is a non-trivial endeavor for the perceptual-attentional system to accomplish the task. Clearly, a mechanism is needed to ensure that correct identity-location combinations are perceived and maintained.

Several fundamental questions can be asked about the workings of the binding mechanism. How efficiently does it carry out its task across different dynamic input situations? Is it vulnerable to changes in 'where' information but not in 'what' information, or vice versa, or both, or neither? How crucial is the availability of identity information for successful tracking? Does it matter whether the identities are known to the observer (*the role of semantics*)? How susceptible is the binding mechanism to changes in target velocity (*the effect of speed*)? Does the mechanism possess a large parallel capacity so that the observer is able to genuinely track several moving target identities simultaneously, or does it possess only a very limited capacity so that the observer needs to constantly shift the focal attention from one target to another to check out the identities (*the effect of set-size*)?