# MULTIPLE IDENTITY TRACKING AND MOTION EXTRAPOLATION

Ashley Buck and Esa Rantanen
Rochester Institute of Technology, Rochester, NY, USA

Abstract

Multiple Identity Tracking (MIT) is a research paradigm in which individuals track the location and individual identity information of several moving objects in the environment. The present study is an examination of how individuals are able to extrapolate the future movement of moving objects while they are masked. There has been conflicting research on the source of a decline in tracking ability; either the amount of time an object is occluded for, or the distance an object moved during an occlusion. Additionally, previous research has not included the use of a secondary visual search task in a mask. Our design was modeled after a task of a pilot, who has to divide his or her attention between flight information on a head-up display (HUD) and traffic information on a horizontal situation display (HSD), while maintaining good situation awareness on both sources of information.

The purpose of this study was to identify the determinants of performance in tracking multiple moving objects while maintaining their identity-location bindings in the visual short-term memory. This study expanded on past research by investigating the relationship between object displacement during masking of the objects and masking duration (simulating looking away from the HSD) on tracking performance. Isolating and identifying the aspects of tracking multiple objects that are most detrimental to rapid reacquisition of a given target object will help designers, engineers, and researchers identify solutions that would result in improved performance and a lighter cognitive load. No study to our knowledge has examined if poor performance is due to the displacement of an object or the duration of time that passes when a mask occludes the objects in a multiple identity tracking task.

This study also introduced the concept of task switching during an occlusion to an MIT scenario, making the experiment more realistic. Switching attention away from the MIT task may be brief but it requires processing additional information while maintaining the identity-location bindings in memory to quickly reacquire objects to be tracked when attention is again paid to them. Studying performance from this view point yielded results that are more aligned to what could be expected from people engaged in MIT tasks in realistic settings. Finally, this study is unique in that all objects in the task will be potential targets, adding another element of realism to the experiment.

We hypothesized that object displacement would result in poorer performance in identifying targets compared to the duration of the occlusion. In a realistic scenario, this would mean that if a pilot lost track of a moving object when they foveate to the HSD, it would likely be due to the objects moving far from the point they were last attended to, rather than the pilot spending more time on the HSD. Based on the outcomes of previous studies, we expected performance to be best in the lowest object speed, shortest occlusion task condition (smallest object displacement) and worst in the highest object speed, longest occlusion condition (largest object displacement). The results have important implications on estimating human performance in modern aircraft cockpits and for the design of both pilots' tasks and the displays helping them perform the tasks.

## Method

Four combinations of object speeds and occlusion times were examined. These include a fast speed with a short occlusion (resulting in 2.54° visual angle (VA) displacement and 2 s occlusion), fast speed with long occlusion (5.08° VA displacement and 4 s occlusion), slow speed with short occlusion condition (5.08° VA displacement and 2 s occlusion), and slow speed with long occlusion (10.08° VA displacement and 2 s occlusion). The experimental conditions are presented in Table 1 below. We hypothesized that object displacement during the mask would drive performance in identifying targets rather than duration of the mask, based on the outcomes of previous studies (Keane & Pylyshyn, 2006).

Table 1
*Object Displacement Calculations, Which Are Based on a 22.5-Inch PVD at a 500 mm Viewing Distance*

| Condition | Occl. Time | Obj. Spd (mm/s, | °VA/s) | Displ. (°VA) |
|---|---|---|---|---|
| Slow Speed, Short Occlusion | 2 s | 11.1 | 1.27 | 2.55 |
| Slow Speed, Long Occlusion | 4 s | 11.1 | 1.27 | 5.08 |
| Fast Speed, Short Occlusion | 2 s | 22.2 | 2.54 | 5.08 |
| Fast Speed, Long Occlusion | 4 s | 22.2 | 2.52 | 10.08 |

### Participants

Ten participants were recruited to pilot the study from Rochester Institute of Technology's undergraduate student population. All participants had normal or corrected-to-normal vision. All procedures were approved by Rochester Institute of Technology's Institutional Review Board.

### Apparatus and Materials

Stimuli were shown in a program developed in Javascript and run in Java Runtime Environment. A Google satellite image (dark green forest) was set as the background for the HSD and the MIT task. The frame rate was set to 60 fps. Five small aircraft symbols were used as tracking stimuli. An alphanumerical call sign below each object, written in

12 point font, served as each object's identity. Each object moved in square paths at a predetermined speed of 11.1 mm/s or 22.2 mm/s. The objects moved for 7 s per trial before being masked and continued to move behind the mask.

An image of a head-up display (HUD) showing altitude and speed was used as the mask. The participants were tasked with determining if either the altitude or speed was safe (above 500 feet for altitude and below 1000 kts for velocity). The combination of altitudes and velocities was unique to each trial. Participants were prompted in the mask image to check either their altitude and velocity, and had 2 or 4 seconds to make a verbal "yes" (meaning safe) or "no" (meaning unsafe) response. Participants were given a reference sheet to use as needed and they practiced the associations as many times as they wanted to before performing the experimental trials.

Once the mask was removed, the objects froze in place, and the object identities were masked under black boxes. Participants were prompted to click on a specific object using a small pop-up window on the top left corner of the screen. Putting the cursor on an object revealed its identity. Once the participant clicked on the correct object, the objects resumed movement and the pop-up window disappeared.

## Independent Variables

Object speed and mask duration were manipulated. The faster an object and the longer the mask duration, the larger the displacement during the mask. Object speeds were chosen to result in equal displacement in 2 or 4 seconds, to allow comparable conditions to determine if displacement from original position or occlusion time has a greater effect.

## Dependent Variables

Response time, number object identity checks before clicking on the target object, and responses to the mask scenario were measured.

## Design

A factorial 2 $\times$ 2 within-subjects design was used to compare 2 object speeds (11.1 mm/s and 22.2 mm/s) and 2 occlusion times (2 s and 4 s). A within-subjects design of this experiment accounted for individual differences in response time.

## Procedure

All participants were seated 50 cm from a 22.5-in computer monitor. Participants were given a printed reference sheet and the researcher reviewed the instructions with the participants. The reference sheet included a simple representation of the mask scenario to help participants learn where to locate the altitude and speed on the HUD and what was considered "safe".

Participants completed 15 practice trials, or repeated the practice trials as many times as they wished until they felt comfortable with the task, followed by 100 experimental trials. The reference sheet remained in front of the participant during the experiment allowing participants to refresh their memory.

The experimental trials took approximately 20 minutes to complete. Five objects appeared on the screen, within a constraint of at least 1 degree away from the edge of the screen, and at least 1 degree apart. All objects on the screen moved at a consistent speed through the entire experiment, and all objects could potentially be targets (i.e., there were no distractors). The use of distractors in MIT tasks is not needed, because each object's identity is unique and distinct from all other objects (Oksama & Hyönä, 2008).

The participants tracked the moving objects for 7 s after which the mask screen occluded the entire tracking screen; the objects continued to move in the background. The mask appeared for either 2 or 4 seconds in a random order. Participants were asked to respond verbally if the altitude or speed of the object was "safe". Once the mask was removed, participants were presented with the tracking screen again, with the objects frozen in place and their identities masked. Participants were prompted to click on a target object as quickly as possible (Fig. 1). The software used to run the program logged responses and how many times the participant revealed an object identity by running their mouse over the label. Responses to the mask scenario were collected by the researcher by hand.


*Figure 1*. The experimental tasks and procedure. Participants tracked 5 moving objects with unique identities for 7 s (left) until the view of the object was blocked by a mask depicting a head-up display for 2 or 4 s, with the objects continuing to move on the background (center). Participants were required to see if the altitude or speed displayed was safe. After the mask was removed, the object reappeared frozen and with their identities masked. The participants were required to click on a target object queried in the pop-up box on the upper left corner (right)

## Results

Data analysis was conducted using MS Excel and Minitab 18 software. The overall average response time to identify an object was 4.20 seconds (SD = 2.79). Table 2 shows these results by condition. A repeated measures ANOVA was performed; the differences in response times between conditions were not statistically significant, $F(3, 997) = 1.39$, $p = 0.246$, with $R^2 = -.42\%$.

Object label reveals were counted. If a participant only had one object reveal, this indicated that the participant knew the location of the object and this was considered to be perfect performance. Two or more means the participant revealed multiple object labels to be able to identify the correct object (Table 3).

Table 2
*Average Response Times Per Condition, and by Overall Performance.*

| Condition | Average RT (s) | SD |
|---|---|---|
| Overall performance | 4.20 | 2.79 |
| Slow speed, short occlusion | 4.44 | 2.94 |
| Slow speed, long occlusion | 4.16 | 2.46 |
| Fast speed, short occlusion | 3.94 | 2.29 |
| Fast speed, long occlusion | 4.19 | 3.18 |

Participants had perfect performance in about 50% of the trials per condition, and in 54% of the trials overall. Participants had two or less object reveals in 75% of the total trials, and participants revealed three or more objects in 25% of the trials.

Due to the nature of the pilot test, the performance on the task-switching scenario was monitored by two researchers and spot checked for performance. The researchers did not observe any errors in performance during the course of the study, indicating that participants were performing the task-switching task accurately.

Table 3
*Percentage of Object Label Reveals Per Condition and Overall Performance.*

| Condition | 1 rev. | 2 revs. | 3 revs. | 4 revs. | 5 revs. |
|---|---|---|---|---|---|
| Overall performance | 54% | 21% | 12% | 7% | 5% |
| Slow speed, short duration | 53% | 24% | 13% | 5% | 5% |
| Slow speed, long duration | 55% | 18% | 12% | 8% | 6% |
| Fast speed, short duration | 60% | 19% | 11% | 5% | 5% |
| Fast speed, long duration | 49% | 23% | 11% | 11% | 5% |

## Discussion

The response times did not differ between the experimental conditions. There are several possible explanations for this finding. Participants who were unsure about the location of an object could have quickly uncovered the object labels nearby until the object was found (also meaning they had a quick visual search time). It is also possible that response time was not fully representative of performance in this task due to the high number of participants who were accurate on the first or second try (based on object label reveals). These findings suggest that participants were able to perform the task quite well without occlusion time or object displacement impacting performance.

Object label reveals indicated that the task was somewhat difficult for participants, but not impossible. Participants selected the correct item on the first try 54% of the time, and revealed two or less object labels 75% of the time. This suggests that participants generally could keep track of the objects. Revealing more than one object would indicate that the participant did not know where the object was located; choosing the correct object on the second try might suggest that participants had a general idea of where the target object was located, but confused objects that were in a close vicinity to one another.

Previous studies on the ability to extrapolate motion through occlusions have used shorter occlusion times that were less than half of the duration than the present study (Cohen, Pinto, Howe, & Horowitz, 2011; Fencsik, Klieger, & Horowitz, 2007; Franconeri, Pylyshyn, & Scholl, 2012; Keane & Pylyshyn, 2006; Zhong, Ma, Wilson, Liu, & Flombaum, 2014). The findings suggest that it is possible to track multiple moving objects and maintain identity-location bindings for longer periods than previously determined.

## Conclusion

The present study was an examination of how object occlusions and object speeds impact tracking performance when the objects have unique identities. Additionally, the use of a task-switching scenario challenged participants to perform a second task during the occlusions. The results suggested that participants could perform a tracking task of this degree of difficulty, but participants had equal performances in each condition. The implication of these findings is that individuals have a limited ability to maintain the identity-location bindings of objects in their visual short term memory and switch to a brief alternative task without losing these bindings. The findings in the present study do not elucidate if tracking ability in a MIT task is impacted more by object occlusions and object speeds. This question should be further investigated, because knowing the limitations of performance in MIT tasks will aid in the design of appropriate systems in operational settings in the field of aviation.

## References

Cohen, M. A., Pinto, Y., Howe, P. D. L., & Horowitz, T. S. (2011). The what-where trade-off in multiple-identity tracking. *Attention, Perception, and Psychophysics*, *73*, 1422-1434.

Fencsik, D. E., Klieger, S. B., & Horowitz, T. S. (2007). The role of location and motion information in the tracking and recovery of moving objects. *Perception Psychophysics*, *69*, 567–577.

Franconeri, S. L., Pylyshyn, Z. W., & Scholl, B. J. (2012). A simple proximity heuristic allows tracking of multiple objects through occlusion. *Attention, Perception, Psychophysics*, *74*, 691–702.

Keane, B. P., & Pylyshyn, Z. W. (2006). Is motion extrapolation employed in multiple object tracking? tracking as a low-level, non-predictive function. *Cognitive Psychology*, *52*, 346–368.

Oksama, L., & Hyönä, J. (2008). Dynamic binding of identity and location information: A serial model of multiple identity tracking. *Cognitive psychology*, *56*, 237–283.

Zhong, S.-h., Ma, Z., Wilson, C., Liu, Y., & Flombaum, J. I. (2014). Why do people appear not to extrapolate trajectories during multiple object tracking? a computational investigation. *Journal of Vision*, *14*, 1–30.