2015

Learning of Location-Identity Bindings: Development of Level 1 Situation Awareness in an Air Traffic Control-Like Task

Alexander Nalbandian

Esa Rantanen

LEARNING OF LOCATION-IDENTITY BINDINGS: DEVELOPMENT OF LEVEL 1 SITUATION AWARENESS IN AN AIR TRAFFIC CONTROL-LIKE TASK

Alexander Nalbandian and Esa Rantanen
Rochester Institute of Technology, Rochester, NY

Knowing "what is where" is essential to human perception and performance. This knowledge corresponds to the concept of Situation Awareness (SA), specifically Level 1 SA. The underlying research paradigm concerns tracking of identical objects moving on a screen (Multiple Object Tracking, MOT). This method has been useful to investigate the fundamentals of visual tracking, but it lacks a connection to real-world scenarios. In another paradigm, objects tracked have unique identities (Multiple Identity Tracking, MIT) requiring a combination of peripheral and focal perception in tracking. This model has been used to examine air traffic controllers' SA. This paper will report results from an experiment where objects with identities similar to typical air traffic control (ATC) call signs were displayed on a plan-view display. The 4- and 8-object conditions replicated previous research and a 12-object condition simulated the density of typical ATC displays. The displays were static to examine the time required to create identity-location bindings (Level 1 SA). The displays were periodically blanked and participants queried about locations of target objects. Location errors and response times were recorded. The object array size of 4 had the highest accuracy and shortest time to acquire good SA. As the object set size increased from 4 objects to 12 objects, location errors increased dramatically. In sum, participants could reliably retain locations and identities of only about three objects. This finding has implications to the capacity of controllers to maintain adequate SA in future ATC systems.

This research is premised on two main ideas, one applied, and the other theoretical. The applied premise is the need to accurately model air traffic controllers' performance. The primary task of air traffic controllers is to ensure that aircraft within their areas of responsibility maintain at least the minimum standard vertical and horizontal separation from each other at all times. To perform this task, controllers must at all times know the current positions of aircraft they are responsible for as well as their trajectories to predict their future locations. Controllers refer to their ability to track multiple aircraft as a mental "picture". In the human factors research literature an analogous—if not identical—construct to the "picture" is situation awareness (SA). SA is typically defined to have three distinct levels (Endsley, 1995a). Level 1 refers to the perception of elements in an environment, such as aircraft call signs and their spatial locations on the controller's display, Level 2 pertains to the meaning of those elements, and Level 3 to controllers' ability to predict their future states.

With the introduction on the NextGen technology in ATC in the near future, controllers' tasks will shift from active control of air traffic to monitoring of automation (Durso & Manning, 2008; Metzger & Parasurman, 2001). At the same time, the number of aircraft under individual controllers' responsibility at any given time will likely increase. Under such conditions, acquisition and maintenance of Level 1 SA in not a trivial task. Niessen and Eyferth (2001) demonstrated a cognitive model (MoFi) of controllers' SA that posits a monitoring cycle of the radar screen to acquire and refresh all aircraft information, suggesting a continual cycle of learning and forgetting. Forgetting about aircraft is a common error in ATC (Shorrock, 2005), as are perceptual errors of misidentification of aircraft based on both auditory and visual input (Shorrock, 2007). Such errors can be attributed to an inadequate Level 1 SA.

Research Paradigms

The second, theoretical, premise of our work concerns the choice of an appropriate research paradigm for examining human performance in tasks that are relevant to and closely resemble those of air traffic controllers. The multiple (moving) object- or identity tracking (MOT/MIT) paradigm is plainly analogous with ATC tasks. In previous research on MOT it has been constantly argued that the mechanism behind tracking is parallel in nature. The size of the parallel capacity is most explicitly defined by Pylyshyn and Storm (1988), who posited 4-5 visual indexes that move automatically along with the moving objects, as if these pointers were glued to the tracked objects. Very recently, however, this view has been questioned. Oksama and Hyönä (2004) have provided evidence that the efficient tracking performance is based on continuous attention switching between the tracked targets with the help of visuospatial working memory. To study dynamic identity tracking or binding, i.e., the observer's awareness of the location and identity of visual elements at any given time, Oksama and Hyönä (2004) developed a new multiple identity tracking paradigm, which is plainly analogous with ATC tasks. They found that multiple identity tracking performance deteriorated linearly as a function of set-size, tracking time, object speed, and target familiarity. Furthermore, they found that tasks measuring visuospatial memory and attention switching proved to be significant predictors of multiple object tracking performance. These findings are not consistent with the notion of parallel tracking, which is assumed to carry out the task efficiently and automatically without recourse to featural or semantic properties of the objects. On the other hand, they are more consistent with a higher-order post-attentive serial switching account, which operates with the help of temporary memory buffer(s).

Hope, Rantanen, and Oksama (2010) incorporated the concept of entropy, or the magnitude of direction changes in object trajectory, and traditional aircraft call signs to produce an experimental paradigm that could test how well a participant could track moving objects on a PVD. The experimental design was modified from a moving identity tracking (MIT) task, but instead of tracking a set of predetermined objects, all object were potential targets and ATC call signs were used in place of the line drawings or faces. The call signs were so small that the participants had to foveate on each identity. After a viewing period of 20 seconds the objects were stopped and masked and participants prompted to click on a particular call sign. The participant then chose the object in question from the stationary masked call signs. The participant would answer by the clicking on the call sign he thought was the correct one. Independent variables were the number of objects on the screen (4, 9, 14, 19 identities) and the amount of entropy (0.00 meaning a straight line, and 0.69 and 1.00 resulting in constantly changing direction).

A main effect for the number of objects was found. More importantly, a small effect for entropy was found but in opposite direction than hypothesized. Better performance was achieved in higher entropy conditions than when the objects moved along straight lines. Hope, Rantanen, and Oksama (2010) suggested that the level of SA was not at the level 3 that has been suggested by previous literature (Endsley & Garland, 2000) but at level 1 SA. In other words, it appears that the participants did not encode the trajectory information of the moving objects but looked for them at their last known positions. These results suggest that maintenance of SA in ATC is a static visual search task rather than a tracking task.

ATC Display Considerations

The above conclusion is further supported by examination of the speeds symbols of aircraft typically move on ATC displays. Although aircraft move through the airspace at high speeds and control decisions have to be made and actions taken within temporal "windows of opportunity" (Rantanen, 2009), at a perceptual level it may be argued that the movements of aircraft symbols on plan view displays (PVDs) represent a static rather than dynamic situation. To determine the speed of a moving object (e.g., aircraft symbol) on a PVD in terms of degrees of visual angle per second (Deg. VA/s), four parameters need to be known: (1) the ground speed of the object displayed, and in the case of airborne displays, the ground speed of the ownship, (2) the scale of the display, or the area depicted on the display, (3) the actual size of the display, and (4) the viewing distance.

In ATC applications all of the aforementioned parameters vary widely. Controllers can sit back or lean forward as they view their displays, their viewing distance ranging from as much as 1 m (or 1,000 mm) to as little as 250 mm. For our calculations here, the typical viewing distance of computer displays in experimental conditions of about 500 mm seems to be a reasonable average of also the viewing distances of ATC PVDs, and hence we will use this value in subsequent calculations.

The second source of variability is the scale of the PVD, which can be freely selected by controllers from a wide range of values. There are no hard limits for display range, but in practice these typically vary from about 200 nm across the display for large enroute sectors with little traffic to about 20 nm across the display in approach control operations. Finally, aircraft ground speeds vary from supersonic for military jets (since the retirement of the Concorde) to about automobile highway speeds for small aircraft. Hence, we can estimate the range of aircraft true air speeds (which equals their ground speed in the absence of wind) to be from 1,150 kts (Mach 2 near tropopause) to 80 kts.

Given the display range options and the range of aircraft speeds as described above, and assuming a modern 20 x 20 in PVD, we can calculate the range of target (or object, i.e., aircraft position symbol) velocities on ATC PVDs in terms of VA to be from 0.25 deg/s for a military jet maneuvering at about Mach 1.2 at low level and viewed on a PVD with a range of 50 nm to 0.02 deg/s for a general aviation on a cross-country flight at 100 kts, viewed on a display showing 100 nm across. In Table 1 some typical target velocity values on PVDs have been calculated. Given these speeds and the perceptual cycle of controllers monitoring traffic on their displays and continually updating each aircraft's position information, we argue that the displayed information is more static than dynamic in nature. Futhermore, contrasting the velocities in Table 1 to those used in experimental MIT research (2.6

to 10.7 DegVA/s in Oksama & Hyönä, 2008, and 4.32 DegVA/s in Hope et al., 2010) casts doubt on the usefulness of the MIT paradigm to study air traffic controllers' performance in operational settings.

Table 1.
Some Typical Values of Object Speeds on PVDs in ATC. The Calculations Are Based on a 20-in Diameter PVD Viewed at a Distance of 500 mm.

Type of flight	Target GS (kts)	PVD range (nm)	Tgt speed (mm/s)	Tgt speed (DegVA/s)
Fighter jet, maneuver	790	50	2.23	0.26
Airliner, approach	230	50	0.65	0.07
Fighter jet, cruise	1150	200	0.81	0.09
Airliner, cruise	550	200	0.39	0.04
GA, cruise	100	80	0.18	0.02

Performance Measurement Considerations

Hope, Rantanen, and Oksama (2010) masked the object identities on the display when their participants were asked to click on a given identity. The participants could move the cursor on any masked identity which was unmasked, allowing for verification of the identity. If the first guess was wrong, the participants checked another masked identity. In this paradigm response accuracy could be expected to be 100% (i.e., the correct identity could eventually be found) and performance was measured by response time. According to this paradigm, if the participant has good Level 1 SA and knows where the queried object is, he or se will click on the right (masked) object the first time. If the participant's SA is poor, he or she must check multiple objects before finding the right one, which is reflected in an increasing response time (RT), which should be a multiple of the number of objects checked before finding the queried one.

The above paradigm is primarily concerned with identity error, or confusion between identities of displayed objects. It does not measure location error, as the objects remain visible (albeit with masked identities) throughout the experiment. Another paradigm may be considered to also measure location error, that is, uncertainty of the actual location of the queried object. To accomplish this, the objects on the screen could be masked completely, that is, the display is blanked upon query of a particular object. Now the participant must not only recall the approximate location of the queried object, but click on the exact location from memory. Identity errors would be manifested by clicks closer to the location of a wrong object than the queried one. Finally, instead of examining performance in a snapshot fashion, it may be of interest to do a timeline analysis of multiple trials. Such analyses would reveal learning of location-identity bindings over time.

Method

Participants

A total 45 participants were recruited from the student population at Rochester Institute of Technology. Most participants had normal or corrected to normal vision. Six participants were removed from the data because of difficulties completing the task.

Apparatus

A MacbookPro laptop computer was used to run the experiment. The computer had a 15" LCD screen to display the experimental objects. Screen resolution was 1024 × 768 pixels. A standard Microsoft desktop mouse was used to move the cursor. The PEBL programming language was used to create the experimental visualization and to collect the data.

Stimuli

The stimuli were objects that mimicked ATC call signs and consisted of 3 letters and 4 numbers. The call sign list was predetermined before set of trials. The objects subtended 2-3 degrees of visual angle and required the participant to foveate on each object to resolve their unique identities. The locations of the objects on the display were random and configured before the experiment. To achieve an appropriate level of difficulty of the objects' identities, same letters or numbers were used at the first, last, and center positions in each of the call signs in various sets (Tydgat & Grainger, 2009). For the remainder of positions the order of the letters and numbers were shuffled to create unique identities for each object. Therefore, in a given set of objects, the same three letters were used and the same four numbers were used to make the identities sufficiently confusable and to make the participants to memorize more than just a single letter from each identity.

Independent Variables

The number of the objects varied between trials (4, 8, 12 identities). These set of objects were used to replicate previous research (Pylyshyn & Storm, 1988; Oksama & Hyönä, 2004, 2008; Alvarez and Franconeri, 2007) and to provide a semi-realistic simulation of the density of ATC displays. Each of the object sets was displayed stationary to examine the study time required to effectively create identity-location bindings.

Dependent Variables

The central dependent variable for this experiment was the error between the location of the object in question and the participant's response (a mouse-click on the display). The error was measured as the distance between where the participant clicked and the location of the queried object. Additionally response time, measured from the end of the study period to the mouse click on the queried object, was recorded.

Experimental Task

The task of the participant was to study the entire set of objects throughout the display for a period of time. Based on a pilot study, the study time was 0.7 seconds per object (or 0.7 s × number of objects). After that the screen went blank and the participant heard an audio clip of a particular object identity. Simultaneously, the mouse cursor reappeared at the center of the display and the participant then clicked on the location the queried object. This process was be repeated for 50 trials for each set of objects.

Procedure

The experimenter explained the task to the participant and answered any questions about the experiment to them. The participant then read through and filled out an informed consent form. Upon completion of the form, the participant was situated in front of the display. The participant was reminded to study all the objects on the display and to click on the location of a particular object upon hearing an object identity on an audio clip. Before the experiment, 5 practice trials were used to help the participant get used to the system. The participant was offered breaks after each set of trials. The order of the set of objects was randomized to minimize any order or fatigue effect. Once the trials were completed, the participant was asked about any strategies used and if they were satisfied by the amount of study time during the debriefing. Participants received extra credit in a psychology course for completing the study.

Results

There was a total of 150 trials for each participant, which adds to a total of 5,850 data records. In the initial analysis of the data, it was determined that some of the response times occurred before any time to distinguish the object identities were possible (below a threshold of 2,777 ms). A Pearson correlation was used to examine response time and accuracy (error measure) tradeoffs for each object set size. Only very weak correlations were found ($r = 0.13$, 0.08, and 0.11 for the 4, 8, and 12 objects conditios, repectively; all $p < 0.01$). If the participants took longer to respond, their accuracy improved only marginally.

The object with the shortest Euclidean distance away from the click location was considered the intended target. By matching the intended target of the user and the queried object, the identity error could be determined (i.e. match or no match). If the intended target matched the queried object, the distance was the location error.

Plots of accuracy versus trial number suggested learning curves . The power law equation ($y = ax^b + c$), was fitted to the data (Newell & Rosenbloom, 1981). The results are depicted in Figure 1. It is apparent that little learning took place after about 10 trials, or that after studying the objects for about half a minute (4 objects), or one minute (8 objects), or minute and a half (12 objects), memory of their locations and identities improved only incrementally. Most importantly, the location errors for 8 and 12 objects remained quite large even after all 50 trials. This finding appears to support our hypothesis that performance in tracking multiple moving identities suffers primarily from poor Level 1 SA.

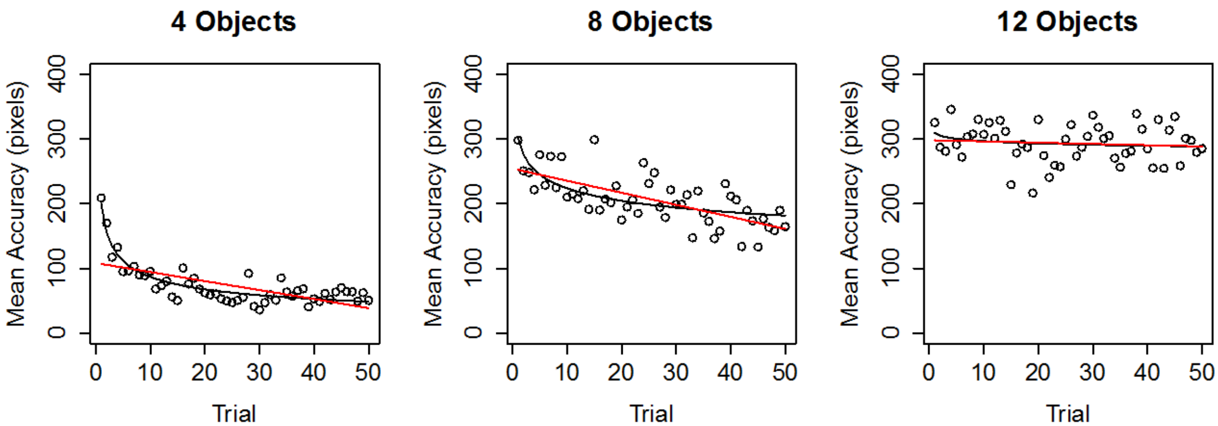

Figure 1. Learning curves of identity-location bindings in an experimental task involving 4, 8, or 12 objects with unique identities on a display. Performance even after 4.6 or 7 minutes of study time for 8 and 12 objects, respectively, was remarkably poor. N = 39.

Learning curves that emerge from average performance of multiple individuals are famously misleading and often mask individual differences and true effects of learning. Therefore, we examined performance of individual participants in each object set size condition. This examination revealed that the participants' performance in the 4-objects condition was generally good and reasonably high accuracy was achieved too quickly for learning effects to emerge. In the 8- and 12-object conditions, however, the performance was much worse and also much too variable to suggest any learning. Because the power law equation fitted very poorly to individual participants' data, no meaningful further analyses (functional data analysis) could be performed.

A holistic view of the results is provided in Table 2. The generally very poor performance is noteworthy. Even in the easiest condition with only 4 objects to memorize over 10% of the objects were misidentified. The means and ranges of the location errors should be judged relative to the display size of 1024 × 768 pixels. Using the percentage of correctly identified objects and the number of objects, it was calculated that only about 3 objects' could be retained in memory during the experiment.

Table 2.
Mean Location Error, Location Error Range, and Identity Error Percentage for each Condition.

Objects	M Location Error (pixels)	Error Range (pixels)	Ident. Error (%)	Approx No. Recalled Objects
4	38.17	1-188	88.24	3.53
8	40.68	2-194	41.11	3.29
12	50.27	3-151	23.37	2.80

Note. The mean location error is only for the correctly identified targets, not for all data points.

Discussion

Air traffic controllers tasks are quite complex and modeling controller performance therefore at least equally complex. Even reduction of controllers' tasks to experimental paradigms such as multiple identity tracking (Hope, Rantanen, and Oksama, 2010) or learning of initial location-identity bindings of several objects on a display (this study) have proved to involve myriad poorly understood variables. Our results also seem to suggest that many experiments on multiple identity tracking may have provided insufficient time for the participants to acquire adequate Level 1 SA for good performance on tracking moving objects, confounding Level 1 SA and tracking performance.

The debriefing questions provided insight into strategies and the study time for the experiment. Participants described using spatial patterns of the objects, chunking the objects identities (Cowan, 2001), and splitting the screen in half to determine which object was where (Alvarez & Cavanagh, 2004). As for the study time, most participants felt they had sufficient study time for the 4 object condition, but inadequate time for the larger set sizes.

The average inspection time of 0.7 seconds per object was not sufficient for higher levels of object set sizes, apparent in the very poor accuracy with 8 and 12 objects. Either learning requires much longer time than this experiment provided, or, more likely, controllers are not committing all the information (object identity and location) to memory, but continually search for the aircraft they need to attend (Hope et al., 2010).

As is usually the case in research of ATC and controller performance using naïve college students as participants, this study, too, suffers from some serious limitations. The "callsigns" or identities of the experimental objects used in our experiment had no meaning whatsoever to the participants. In ATC, specific callsigns have meaning to the controllers, signifying type of the flight (corporate or private aircraft vs. airliners) and possibly much flight plan-related information, such as route and altitude. This information certainly is helpful in maintenance of Level 1 SA. Confusability of realistic callsigns is also not uniform and estimation of appropriate—or realistic—levels of confusability is therefor difficult. Finally, because of the number of variables influencing performance even in simple laboratory experiments, a single study does not allow drawing of many conclusions. For that, series of experiments are necessary. Nevertheless, methodological issues are also important, and we hope that this study makes a contribution in that regard.

References

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science, 15*(2), 106–11.

Alvarez, G. A., & Franconeri, S. L. (2007). How many objects can you track? Evidence for a resource-limited attentive tracking mechanism. *Journal of Vision, 7*, 1-10.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*(1), 87–114.

Durso, F. T., & Manning, C. A. (2008). Air traffic control. In C. M. Carswell (ed.) *Reviews of Human Factors and Ergonomics, 4* (pp. 195-244). HFES.

Endsley, M. R. (1995a). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*(1), 32-64.

Endsley, M. R. (1995b). Measurement of situation awareness in dynamic systems. *Human Factors, 37*(1), 65-84.

Endsley, M. R., & Garland, D. J. (2000). *Situation awareness: Analysis and measurement*. CRC Press.

Hope, R. M., Rantanen, E. M., & Oksama, L. (2010). Multiple identity tracking and entropy in an ATC-like task. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 54*(13), 1012-1016.

Metzger, U., & Parasuraman, R. (2001). The role of the air traffic controller in future air traffic management: An empirical study of active control versus passive monitoring. *Human Factors*, *43*(4), 519-528.

Mogford, R. H. (1997). Mental models and situation awareness in air traffic control. *The International Journal of Aviation Psychology, 7*(4), 331-341.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1-55). Lawrence Erlbaum.

Niessen, C., & Eyferth, K. (2001). A model of the air traffic controller's picture. *Safety Science, 37*(2), 187-202.

Oksama, L, & Hyönä, J. (2004). Is multiple object tracking carried out automatically by an early vision mechanism independent of higher order cognition? An individual difference approach. *Visual Cognition, 11*(5), 631-671.

Oksama, L., & Hyönä, J. (2008). Dynamic binding of identity and location information: A serial model of multiple identity tracking. *Cognitive Psychology, 56*(4), 237-283.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spatial Vision, 3*(3), 179-197.

Rantanen, E. M. (2009). Measures of temporal awareness in air traffic control. *Proceedings of the 53rd Annual Meeting of the Human Factors and Ergonomics Society* (pp. 6–10). Santa Monica, CA: HFES

Tydgat, I., & Grainger, J. (2009). Serial position effects in the identification of letters, digits, and symbols. *Journal of Experimental Psychology: Human Perception and Performance, 35*(2), 480-98.