Modeling of Visual Search and Influence of Item Similarity

Stefan Lindner (stefan.lindner@campus.tu-berlin.de), Nele Russwinkel (nele.russwinkel@tu-berlin.de), Lennart Arlt, Max Neufeld, Lukas Schattenhofer
Department of Cognitive Modeling in dynamic Human-Machine Systems, TU Berlin, Marchstr. 23
10587 Berlin, Germany

Abstract

A modeling approach addressing visual search in an array of items of differing similarity is introduced. The model is able to capture the effects found in a study that varies target-distractor similarity (low vs. high), distractor-distractor similarity (low vs. high) of icons, target presence (present vs. absent) and the set size (8, 16 or 24 icons). To be able to simulate human visual search in such a task with original ACT-R mechanisms we implemented a hybrid search strategy that combines parallel and serial search. The presented model can provide useful insight for researchers interested in modeling tasks containing visual icon search.

Keywords: visual search; similarity; ACT-R; cognitive modeling.

Introduction

Visual search is a general requirement for everyday tasks. Especially for user interfaces it is crucial to find the right icon/button/menu item quickly to proceed with the task and to reach the actual goal. The challenge is to find the target item amongst several, often similar distractor items. Performance in such tasks changes with the number of items on screen. Two search paradigms are known, determining whether the number of items influences search time or not. In the case that the target is similar to other items, search time typically increases roughly linearly with the set size (e.g. Wolfe, 1994). Here **serial search** takes place because the person has to actively attend one item after the other in a serial manner.

In case the searched item is distinctive from the other items (a yellow item between blue items) the subjective feeling is that the item literally pops out from its surroundings. Here, reaction time will not differ too much between set sizes – a phenomenon called the "pop-out effect". This **parallel search** relies on preattentive processes that take place before attention is actively drawn to specific items. Whenever a single visual basic feature (such as color or form) differentiates the target from other items this quick process can occur.

The interesting case is the overlap between those two pure paradigms, whenever a heterogeneous field of items has to be searched.

Our aim is on the one hand to understand how people cope with such search demands and what kind of strategies they use. On the other hand we want to model such search behavior to be able to predict the usability and search time of interfaces.

The cognitive architecture ACT-R (Anderson et al., 2004, Anderson, 2007) offers a visual module that is able to address both search paradigms and also a module for motor output to enable realistic predictions about reaction times in visual search tasks. The visual module has two subsystems, the **where system** and **what system**. The where system simulates preattentive processes and relies on well accepted theoretical concepts (Wolfe, 1994; Treisman & Gelade, 1980). Each visual item has features such as type (text, or oval for a button or others), color or width. It is possible to search for items with a specific feature. As a response to such a search request a visual location of such an item is returned. In the next step the visual attention can be directed to this location. The first process needs no time, the second process does need time. A shift of visual attention takes 135ms - 50ms for the production to fire that elicits the request of the shift and 85ms for the shift itself.

But how is visual search executed that is neither purely serial nor parallel in nature? Do people use strategies to find their target item quicker within larger distractor sets, and does an inhomogeneous distractor set regarding similar features (e.g. Duncan & Humphreys, 1989) further influence visual search apart from the above mentioned mechanisms?

The main goal of the paper is to explore the possibilities of accurately modeling visual search in environments with objects of differing similarity in the cognitive architecture ACT-R.

A number of ACT-R models exist that address visual search with different variations (Fleetwood & Byrne, 2006; Everett & Byrne, 2004). Fleetwood and Byrne manipulated set size and quality of icons in a computer-based target identification task. Icon quality was realized by the level of distinctiveness and complexity of icons. Good quality icons were easily distinguishable from others (on a preattentive level). Evidence in the eye tracking data showed that users were able to preattentively discriminate subsets of visual objects in conjunction search tasks, but here the number of similar items were held constant. Fleetwood and Byrne built two ACT-R models to simulate experimental results and managed to achieve a good fit.

There are also a number of ACT-R modules that aim at a more fine-grained modeling of certain aspects of visual cognition. The EMMA-module (Eye Movements and Movements of Attention; Salvucci, 2001) attempts to better model the intricate relationship between eye movements and

The cognitive processes that closely interact with them, while the PAAV module (Nyamsuren & Taatgen, 2012) allows for the incorporation of bottom-up processes. Our model, however, does not make use of any specific bottom up-processes of visual search. Our rational for that is two-fold. On the one hand, owing to the specific structure of the experiment, top-down search of the target item is generally encouraged and then reinforced through practice.

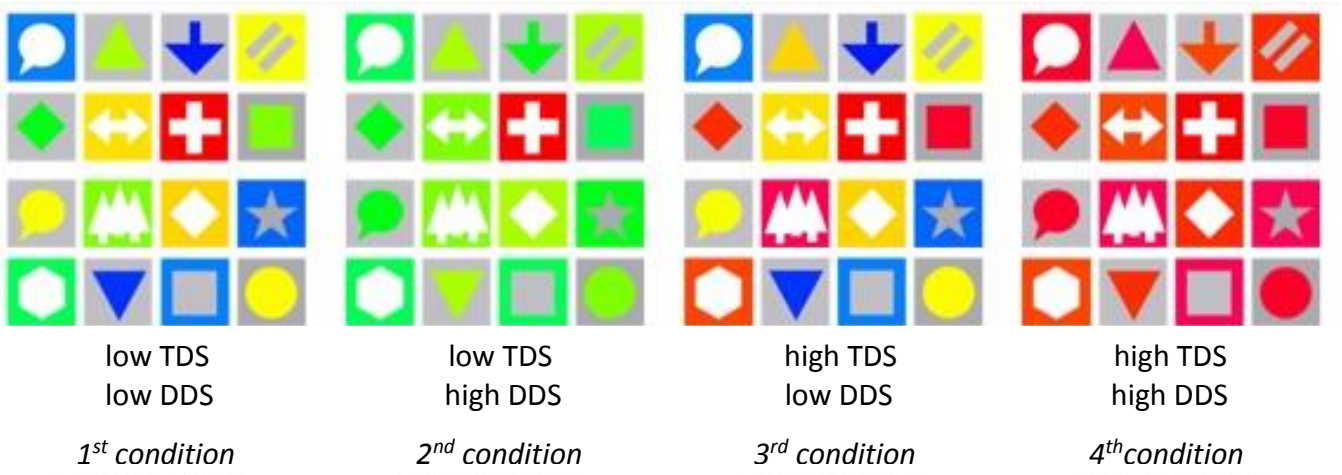

Figure 1: Experimental similarity conditions according to color. Target is the white cross on red ground. (for demonstration; not original icons used)

On the other hand, more importantly even, we are interested in the possibility to model visual search with the core ACT-R mechanisms. While a very fine grained modeling of visual processes has its place, for most task models - especially if they are not primarily focused on the visual aspect of the task – it may be much more realistic and efficient to use a simple model that captures the general behavior reasonably well.

To that end we took an experiment conducted by Trapp & Wienrich (2017) that looks at visual item search independence of four factors: Target-Distractor similarity (TDS), similarity between distractors (DDS), the presence of a target (target presence) and the overall amount of icons present (set size). The experiment is particularly well suited for modeling attempts. It demands the active consideration of both the absolute and relative properties of visual icons such as location, color and form - and therefore tests ACT-R's modeling capabilities in all of these areas, as well. The variation of set sizes also allows for the isolation of invariable mechanisms and those that are dependent on the size of the visual search area.

After presenting the original experiment and its main findings the modeling approach will be introduced. We will first describe the basic model in ACT-R and then move into specific modifications that allowed the final model to capture the experimental results well. To be maximally instructive to future modelers of similar visual mechanisms, we will also shortly discuss several modeling dead ends.

Experiment

The two main independent factors in the experiment by Trapp & Wienrich were Target-Distractor similarity (TDS; low vs. high) and Distractor-Distractor similarity (low vs. high) (see Figure 1). Similarity was realized by the color of the icons. Two further independent factors, target presence (present vs. absent) and the set size (8, 16 or 24 icons) were completely crossed with the similarities, resulting in a 2x2x2x3-factorial setup and a total of 24 experimental conditions. Each participant conducted 12 trials of each condition (for a total of 288 trials per participant), constantly switching between conditions in a fixed blocked fashion. The participants performed a visual search task on a 10" mobile touch device, in which they had to find a specific target icon within a set of distracting icons.

Each trial was performed in the following manner: After the target icon was shown for two seconds, a fixation cross was presented in the center of the screen to ensure a standardized gaze point for all participants. After the fixation cross disappeared, a set of icons was shown. When the target icon was present in the set, the participants had to find and select the target icon as fast as possible. Whenever there was no target, they had to select a specific button at the bottom of the screen to indicate the absence of the target icon. Subsequently, they received feedback on whether their answer was true or false. The reaction time was recorded for each trial and served as a performance measurement. The experiment comprised 18 participants in total (11 male and 7 female) aged between 18 to 30 years.

Both main and interaction effects of TDS, DDS, set size and target presence were consistent with the experimenters' predictions and previous findings. Their main findings were as follows (see also figure 3):

1) The first two conditions (both low TDS) produced low reaction times that showed only a very slight increase with set size.

2) The third condition (high TDS and low DDS) produced moderate reaction times and increased with set size.

3) The fourth condition (high TDS and high DDS) produced high reaction times that increased strongly with set size.

4) The absence of the target item increased reaction times only slightly and by a constant term in the first two conditions. In the third and fourth condition the difference strongly increased with set size.

Model

In order to capture these effects first a basic model in ACT-R was created in a way that required the fewest assumptions while still being able to successfully solve the task in all conditions. Instead of icons the model interacted with oval-objects in the Lisp-GUI with corresponding colors. Instead of the graphic on the icon, text codes were used simulating the visual feature that requires attention shifts. Both this basic and the later, modified model were originally created as part of a student project.

Basic Model

At the beginning of each trial, the model starts by encoding and memorizing the target icon in short term working memory (imaginal buffer). When the fixation cross appears, the visual focus is set on it. Starting with the appearance of the icons the model uses a search routine to scan the graphic user interface (GUI) for the target. Using preattentive perception via the *where* system, it starts a visual-location request for the target color. Its visual attention is then directed to such an item location in order to encode it (text code or icon graphic). The current icon and the target icon (stored in working memory) are compared. Whenever the two items match, the icon is selected. If they do not match, the next item with the target color is picked out and the process repeats until all items with the target color have been attended. If there is no unattended item left, the "not present"-icon at the bottom screen is selected.

While this search routine could plausibly simulate human behavior, this first model had several shortcomings. Most problematically, almost all model behavior was longer than the participants'. This difference was most pronounced in conditions 3 and 4 where many distractor items match the target color and thus the "naive" model had to spend a large amount of time on time-costly fixations of the *what*-system. An additional problem was the fact that the model produced shorter reaction times with no target present (compared to the same condition with target present) in the first two conditions. This was mainly due to the additional visual fixation on the target when the target was present.

Model Changes

To increase the speed, while keeping the model psycho-physiologically plausible, we realized three adjustments: The first adjustment was to move the starting position of the cursor to the center of the screen, assuming that most participants would keep the finger in a click-ready position over the display to be able to react faster. Secondly, as soon as the *where*-system returns a new visual location, two processes start in parallel. While the visual attention is drawn to the location, the manual system prepares to start moving the finger towards the new candidate item. This change was implemented to reflect a routine task handling with subjects constantly anticipating and preparing the next step of the task. Thirdly, while the movement towards an icon takes place, the model already starts to prepare the next motor movement (the pressing of the icon). ACT-R allows for this kind of parallel working of the motor module (here specifically via the "preparation: free" command) as long as the different processes are in different stages of the preparation-initiation-execution sequence that makes up all motor processes. Psychologically, this change can be justified by the assumption that most participants are well-versed in the action of pressing an icon on a touch screen. A procedural acquisition of a combined movement by the participants that does not require several separate preparation and initiation phases is therefore plausible.

Hybrid search strategy

The most important change, however, was the remodeling of the general search in a way that it required fewer attentional fixations, driving down reaction times especially in conditions 3 and 4. Since the fixation of every candidate item (i.e. items of the same color as the target item) was not reconcilable with observed reaction times, the next logical step was to use a strategy that searches an entire cluster of candidate items with one fixation. The new search algorithm (figure 2) thus consists of three main steps (3 productions in ACT-R).

1) A preattentive request (where system) is issued for a previously unattended location with the target color and the lowest x and y values (light blue arrow).

2) If the icon does not match the target icon, the model scans the entire row for the target comparing the "width" of the text (black arrow). Width is processed preattentively, and the target had a unique text width, allowing it to be a search criterion. Psychologically, this much faster search assumes that a visual scan within a short row (here 4 items) allows the shape of the target item to visually "pop out" as well (an assumption that we globally allowed only for colors). When the icon is located in the row, it can be found directly.

3) In case of finding nothing, the model jumps to the nearest oval with the correct color below the current row (red arrow).

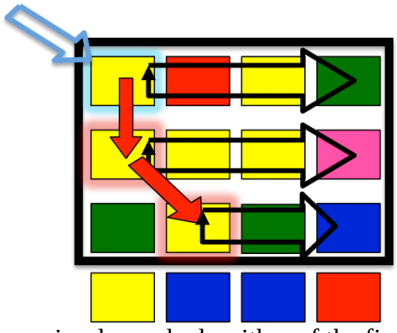

Figure 2: Core visual search algorithm of the final model.

The aggregate of these adjustments allows our model not only to meet the general level of reaction time of the empirical data.

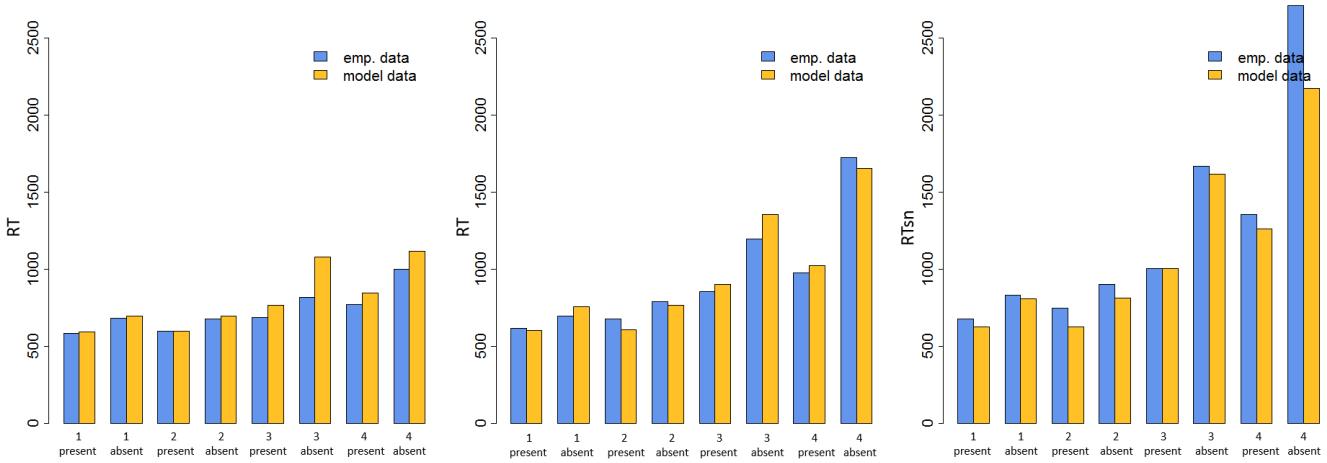

Figure 3: Empirical and Model reaction times for all four principal experimental conditions. Each condition is in turn divided by target present/ not present.

The model now also produced reaction times that are longer in conditions with no target present without any further assumptions.

It should be noted that in this paper, our goal was to recreate reaction times rather than the exact visual paths or fixation patterns. In fact, all models that assume that

1) exactly one fixation per row is needed to scan all candidate items and
2) search is conducted in a structured manner from top to bottom

make the same temporal predictions.

Model Specifications

In the models, no ACT-R parameters were used. The declarative memory consists of a goal chunk and a chunk that stores color, text and width of the current target item. The final model and the GUI are published online at https://depositonce.tu-berlin.de/handle/11303/338. To obtain the simulation results, the model was run 1000 times in each condition.

Results and Model Fit

The following table shows statistical results of the fit. The overall RMSSD (a measure for the absolute distance between model and experimental data; Schunn & Wallach, 2005) of the model is 1.74.

	RMSE	RMSSD	Correlation
Set size 24	102.34 ms	1.34	0.99
Set size 16	172.12 ms	1.20	0.98
Set size 8	179.62 ms	2.42	0.94

Table 1: Statistical model fit (RMSE: absolute fit; RMSSD: absolute fit standardized by the experimental data's standard deviation).

Comparing the empirical data and the model for the set size of 24 icons indicates a good fit over all conditions (figure 3). It captures well the relative trend in all 3 set sizes, both concerning the similarity conditions and the target presence.

The absolute reaction times match reasonably well, although the fit is best for large set sizes. Almost all of the difference between model and experiment results from conditions 3 and 4 when the target is not present. Especially in condition 4 the difference is not within the standard deviation anymore and therefore has a great effect on the RMSSD.

The reaction times reflect the fact that the model is using both the serial and the parallel visual search. While in condition 1 and 2 the target pops out immediately, the model has to use a mixture between parallel and serial search for finding the target fast enough in the other conditions. Despite the fit getting a little less precise in the 3rd and 4th condition the model's searching algorithm seems to be a good approximation to the human visual search behavior.

Discussion

We introduced two modeling approaches. The first one was a simple, reasonable model to address visual search in a task that includes different similarities between target and distractors. This basic model did not capture well the effects found in the experiment. With three adjustments and a new way to describe a mixture of parallel and serial search the new model could capture the empirical data well. The general mechanism used in the model might be helpful to researchers who model visual search in applied tasks, especially for tasks where time is sparse and people try to be as efficient as possible.

To better judge the quality of the current model, it would be useful to compare it to a model that uses both EMMA- and PAAV-module and thus implements more sophisticated mechanisms including those that deal with bottom-up visual processing. Another factor that could also be included in future models is the influence of expectations on visual

search patterns, as described in Lindner & Russwinkel (2016).

Furthermore this account is a theoretical concept that should be tested in subsequent experiments. To test the assumption of the pop out of items on one row, a variation of the current experiment could test participants with the screen presented vertically and horizontally. Another variation could address the question of a possible strategy change when the number of distractors similar to the target increases. This can be done by presenting conditions in which only a small number of distractors is clearly different from the target. All future experiments should involve eye-tracking to better track the attentional focus of participants. This in turn should allow for better model construction and evaluation. Furthermore, we would like to test this visual search concept on our model of learning and unlearning of app usage (Prezenski & Russwinkel, 2016).

Acknowledgments

We would like to thank Anna Trapp and Charlotte Wienrich for sharing their experimental data and Lisa-Madeleine Dörr for creating multiple ACT-R GUIs for our model and for providing modeling ideas. We also thank Lou Conradi and Philipp Wolfgang Klemm for valuable ideas regarding model construction.

References

Anderson, J. R. (2007). How Can Human Mind Occur in the Physical Universe? New York: Oxford University Press.

Anderson, J. R., Bothell, D., Byrne, M. D.,Douglass, S., Lebiere, C., & Quin, Y. (2004). An integrated theory of mind. Psychological Review, 4, 1036–1060.

Duncan, J., & Humphreys, G. W. (1989). Visual Search and Stimulus Similarity. Psychological Review 96 (3), p. 433–58.

Everett, S., P., & Byrne, D.,B., (2004). Unintended Effects: Varying Icon Spacing Changes Users' Visual Search Strategy. In Proceedings of ACM CHI '04: Conference on Human Factors in Computing Systems, Vienna, Austria: ACM.

Fleetwood, M. D., & Byrne, M. D. (2006). Modeling the visual search of displays: a revised ACT-R model of icon search based on eye-tracking data. *Hum.-Comput. Interact.* 21, 2 (May 2008), 153-197. DOI=http://dx.doi.org/10.1207/s15327051hci2102_1

Lindner, S. and Russwinkel, N. (2016). Modeling of proximity-based expectations. In D. Reitter & F. E. Ritter (Eds.), Proceedings of the 14th International Conference on Cognitive Modeling. University Park, PA: Penn State.

Nyamsuren, E. & Taatgen, N. (2012). Pre-attentive and attentive vision module. In N. Rußwinkel, U. Drewitz & H. van Rijn (eds.), Proceedings of the 11th International Conference on Cognitive Modeling, Berlin: Universitaetsverlag der TU Berlin.

Prezenski, S. and Russwinkel, N. (2016). Towards a general model of repeated app usage. In D. Reitter & F. E. Ritter (Eds.), *Proceedings of the 14th International Conference on Cognitive Modeling*. University Park, PA: Penn State.

Reifers, A. L., Schenck, I. N., & Ritter, F. E. (2005). Modeling pre-attentive visual search in ACT-R. In B. Bara, L. Barsalou & M. Bucciarelli (Eds.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates.

Salvucci, D. D. (2001). An integrated model of eye movements and visual encoding. Cognitive Systems Research, 1(4), 201–220.

Schunn, C. D., & Wallach, D. (2005). Evaluating goodness-of-fit in comparison of models to data. Psychologie der Kognition: Reden und Vorträge anlässlich der Emeritierung von Werner Tack, 115-154.

Trapp, A. K., & Wienrich, C. (2017). App icon similarity and its impact on visual search efficiency on mobile touch devices. *Manuscript submitted for publication.*

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97–136.

Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review, 1(2),* 202–238.