The effect of visual cues on workload and learning efficiency in user onboarding

Björn Strömstedt, Frederikke Pape Kofod, Hjalmar Norell, Jakob Bremer, Josef Jönsson, Karoline Engström Poulsen & Olle Bridal

User onboarding is a tool used for introducing novice users to softwares. Based on research on multimedia learning, this study investigated effects of different types of visual cues on users' learning efficiency and workload in context of a user onboarding software tutorial task. The tutorial was constructed based on the software Watchout by Dataton, where participants were asked to solve tasks related to projection-mapping. Experimental participants were randomly assigned to one of three conditions: (1) as baseline, a text-only tutorial; (2) a tutorial with additional shadowing cues; and (3) a tutorial with additional shadowing and gaze cues. An ANOVA and post hoc pairwise t-test was used to compare time to task completion between the three conditions. It was found that the task completion time was significantly lower in the text condition in comparison to the other conditions, suggesting that visual cues in user onboarding can facilitate multimedia learning by offloading working memory in novice users. This finding was further supported by questionnaires related to workload and subjective experience. For future studies, it is suggested to isolate the effect of gaze to investigate it as a main effect, along with using a wider range of eye-tracking measures.

Introduction

The relation between simplicity and complexity has always been a hard bargain. In order to create a software that has great depth you can not always provide the most usable solution. Many modern video editing softwares has taken the road of complexity. Companies have however tried to ease the learning curve for new users. Different kinds of tutorials can be used in order to introduce the software before the user has to make their own decisions. We wanted to investigate this aspect of the user experience further by adding visual cues to three groups of test subjects.

The software used for testing in this study is called Watchout and is made by the company Dataton (Dataton, n.d). It is a widget based video/audio manipulation software that resembles the likes of Adobe Premier Pro and Adobe After Effects (Watchout, n.d). The reason we chose to create a tutorial for Watchout was because it did not have any introductory material built in the software prior to when the study was made. By investigating how people respond to different kinds of introductions we wished to contribute to the field of cognitive science and human-computer interaction.

Background

The study of attention is concerned with where bottlenecks occur and how information is selected. This report is concerned with visual attention, and even though the brain devotes a large part of neural activity to visual processing (Anderson, 2014), it is impossible to process all external information in parallel. Therefore attention is necessary to prioritise a subset from all of the available information for further processing. Because of that, it is important to consider the visual design of the software in relation to goal-directed and stimulus-driven attention. It is important to take the task into consideration and avoid distracting elements that might affect learning and performance.

The Cognitive Theory of Multimedia Learning (CTML) builds upon cognitive psychology and seeks to explain how information is processed and memory is stored in order to facilitate better learning. The theory has three underlying assumptions: there are two separate channels for visual and auditory information processing respectively, the processing has limited capacity, and finally, this is an active process of filtering, selecting, organising, and integrating information (Mayer, 2010). Additionally, Mayer discusses three memory stores involved in information processing: sensory memory which perceives stimuli and stores for a very short time, working memory which is an active processing of the stimuli, and long-term memory that works as information storage for later retrieval. Sensory memory and long-term memory have unlimited capacity for holding information, whereas working memory has limited capacity and therefore is a bottleneck in the system where higher level cognition determines what to attend to, as explained in the earlier paragraph on visual attention (Mayer, 2010).

Cueing effects provide the learner with different types of visual cues to complement the preexisting learning instructions. The technique is commonly approached by presenting the learner with typographical cues such as capitalization or colour variation, or more complex cues, such as summaries or previews. If designed correctly, seen in relation to the format of the learning content, the cueing will direct the learner to the key concepts and stress important information and studies have repeatedly been able to show how cueing effects can be beneficial in the circumstance of learning (e.g. Rickards et al., 1997; Lorch et al., 1995; Glynn & Di Vesta, 1979). Drawing on the *load-reducing method* by Mayer and Moreno (2003), cueing effects can be seen as a framework enabling exclusion of extraneous material along with highlighting relevant material, which will be investigated by introducing the visual cues of *Spotlight/Shadowing* and *Gaze.*

The use of *spotlighting* in onboarding relies on two main objectives; dimming out irrelevant visual stimuli in the interface, referred to as *shadowing,* and highlighting relevant parts of the interface using a *spotlight.* Hence, the implementation of spotlight cues facilitates that the software-user is less exposed to extraneous material as it is either dimmed, blurred, or blackened out and thereby can focus their attention on the important information, which is additionally accentuated by the spotlight.

A gaze cue guides the user's attention to important information by demonstrating a suggested gaze-path for the user to follow. The main function of the gaze-path is to illustrate which part of the screen is important, at what specific time, along with providing the user with a 'route' to follow. The suggested gaze-path may be the most optimal route in terms of time, distance, or other efficiency measures, or it may be inspired by pre-recorded gaze-paths from an 'expert user' who navigate efficiently in the software. The underlying intuition of introducing a gaze cue is based on research indicating differences in attention allocation between novices and experts (Canham & Hegarty, 2010).

The general field of eye tracking technology is concerned with tracking the movements of the eyes and can be approached in various ways. The technology is applicable in numerous domains and has a wide variety of implications (Alemdag & Cagiltay, 2018). The most common measures extracted from eye-tracking data are associated with *fixations,* which refer to the stable state of an eye at one point, being reflective of cognitive processing, and *saccades* which describe the quick eye-movements between fixations and reflect the change in visual attention (van Gog & Jarodzka, 2013).

Aim and hypotheses

This study investigates whether the presence of different visual cues in software tutorials facilitate a more successful user onboarding process and improve the learning efficiency for novice. The cues are selected based on findings in CTML and cueing effects. The following hypotheses are set out for investigation:

H1: Shadowing extraneous information on an interface will reduce the user's cognitive load and decrease time to completion.

H2: Adding an additional gaze that guides the user where to look will further decrease time to completion.

Methodology

Material

The Watchout software was used, along with some media-files to work with, to study task learning. The software ran on a stationary computer which had the eye-tracker SMI RED500 installed. The eye-tracker and its software was used to measure time of completion during the tasks, fixations and saccades. A separate laptop computer was used to present the prototype onboarding interfaces to the participants. One form was used to gather general information about the participants (age, gender, education, technical experience) and another for informed consent.

Two surveys were also incorporated in the study. NASA Task Load Index (NASA-TLX) was used as a tool to measure the test subjects' subjective mental workload assessment of the task (NASA, 2019). User Experience Questionnaire (UEQ) was used as a tool to measure the user experience of the tutorial the test subject was conditioned to (Schrepp, 2019).

Participants

A total of 21 participants (*N* = 21) were assigned one out of the three conditions, having 7 participants in each. The sorting of participants, to each group, was fully randomized. The participant pool consisted of 9 female and 12 male participants aged 21-27 years old (*M* = 23.19, *SD* = 1.83). Ultimately, the data of 3 participants was rejected and discarded. For those three people, the overall quality of the eye-tracking and its linked calibrations were deemed insufficient.

All participants were native Swedish inhabitants and spoke swedish as their first language. The study employed a convenience sampling method to recruit its participants. 15 out of the 21 participants were students of the Cognitive Science program and the rest were other types of students, all studying at Linköping University.

Experimental design and procedure

In the experiment carried out, the participants are randomly assigned to three groups which receive user onboarding interfaces of three different forms with instructions to some fundamental tasks. The first group receives instructions in a visual interface where shadowing is used to blur out redundant information. The second group is instructed by an interface like the one in the first group, but with an added queue to assist eye-gaze. The third and last group is the control group which only receives an interface with verbal instructions, without visual cues. The participants then perform the same tasks in the software. To operationalize the dependent variable of learning efficiency, the measures time of completion, saccades and fixations are recorded when the learned tasks are performed. After completing the tasks, the participants answer questionnaires regarding their subjective experiences of the design of the onboarding interface, as well as perceived workload.

Prototype design

The prototype interfaces were created in the sketching software Figma (Figma, n.d.). It was made by taking images of the software Watchout and merging those with shapes created in Figma. With the use of Figma's "Smart Animate"-tool, the animations and transitions could be made by connecting different frames with each other. The reason transitions between each frame was implemented was because previous research has shown that animacy quality of materials affects visual, attentional and linguistic processes (Pratt et al, 2010). The tasks created in the prototype - and later on performed by the user - were divided into subtasks in order to simplify the instructions for the user. The first task was for the *shadow* condition and *shadow + gaze* condition made up by 10 frames, and for the text condition 4 steps. The second task was separated into 24 frames for the *shadow* and *shadow + gaze* conditions, and 9 steps for the text condition.

Results

There was a statistically significant effect of visual cues given during the task and time to completion, $F(2, 15) = 10.42$, $p = .001$. The Bonferroni post hoc tests revealed that there was a statistically significant decrease in time to completion between *text/shadow* ($p = .004$) and *text/shadow+gaze* ($p = .003$). There was no statistically significant difference between *shadow* and *shadow gaze*.

Mann-Whitney *U* test showed a statistically significant difference between the *shadow + gaze* and *text* conditions in Attractiveness ($U = 9.00$, $p = .034$, $r = .544$), Efficiency ($U = 9.50$, $p = .032$, $r = .523$), Dependability ($U = 7.5$, $p = .017$, $r = .591$) and Perspicuity ($U = 10.5$, $p = .042$, $r = .492$), whilst not in any of the other scales or conditions.

The results from the experiment also revealed that the text-condition yielded a minimum of 8 clues and a maximum of 19. The *shadow* and the *shadow + gaze* conditions instead yielded a minimum of 1 clue and a maximum of 7 respectively. There was a statistically significant difference in the amount of clues given during the duration of the tasks, $F(2, 18) = 22.32$, $p < .001$. The Bonferroni post hoc tests revealed that there was a statistically significant decrease in the amount of clues given between *text/shadow* ($p < .001$) and *text/shadow+gaze* ($p < .001$). There was no statistically significant difference between *shadow* and *shadow + gaze* ($p = 1.000$). Additionally, the analysis revealed a relationship between time to task completion and amount of cues.

The results from a Shapiro-Wilk test was not statistically significant, $W(18) = 0.97$, $p = .786$, thus the data is judged to be normally distributed which validates the use of parametric testing. There was a statistically significant difference in the feeling of frustration during the duration of the tasks, $F(2, 18) = 5.209$, $p = .0164$. The Bonferroni post hoc tests revealed that there was a statistically significant difference in frustration between *text/shadow* ($p < .001$) and *text/shadow+gaze* ($p < .001$). There was no statistically significant difference between *shadow* and *shadow + gaze* ($p = 1.000$).

The results from a Shapiro-Wilk test was not statistically significant, $W(18) = 0.96$, $p = .427$, thus the data is judged to be normally distributed which validates the use of parametric testing. There was a statistically significant difference in the feeling of effort during the duration of the tasks, $F(2, 18) = 3.99$, $p < .001$. The Bonferroni post hoc tests revealed that there was a statistically significant decrease in the feeling of effort between *text/shadow+gaze* ($p = .020$). There was no statistically significant difference between *shadow/ shadow + gaze* ($p = 1.000$) and *text/shadow* ($p = .079$).

Discussion

The results of the study showed a statistically significant difference in time of completion between the text and the text with shadowing conditions, as well as between text and text with shadowing and gaze cue. These results conformed with H1. However, this statistical significance between the difference in time of completion was not found between shadowing and shadowing with the gaze cue. The results therefore indicate a rejection of H2.

Even though a significant difference was found between the *text* conditions and *shadow* and *shadow + gaze* respectively, it is important to remain critical towards the causality of the results. One critical note is that, not only did visual cues differ across conditions, a short text instruction was given along with the *shadow* and *shadow + gaze* condition as well, and this text instruction was not provided in the same manner as the text in the baseline condition. The *text* condition received a notable number of clues compared to the other conditions, indicating that performing tasks under this condition was more difficult. In the baseline, participants received all information at once on how to solve the task, whereas the *shadow* and *shadow + gaze* conditions had ongoing pieces of text for each step towards solving the task following the cue. This difference might be decisive and cause a decreased isolation of the main effects.

In the process of recruiting participants, this study experienced some hindrances as the 2020 Covid-19 situation resulted in the closing of Linköping University, the location of our experimental laboratory. In addition to a general downscaling of schedule, all future decisions had to be taken with an enhanced primary focus on protecting the participant's health and integrity. Furthermore, a shift in sociological behavior could be seen the further the Covid-19 pandemic progressed. A large sum of the students at Linköping University decided to travel to their home town until the end of the semester, while a considerable fluctuation of general social mentally could be seen everywhere; students no longer wanted to leave their home. Ultimately, this meant that the sample size had to be drastically decreased ($N = 51$ reduced to $N = 21$) and recruitment had to be done through convenience sampling.

While much research has been done in the area of multimedia learning and specific visual cues, we acknowledge the fact that there is a very limited amount of relevant prior research done on our specific visual cues; spotlight and gaze. In the exploration of this new area of research, the researchers employed a greater sense of carefulness, designing the research itself. The lack of prior reliable data required us to limit the scope of the study. Additionally, we adapted a stricter framework for what trends or relationships that could be deemed meaningful. That being said, the researchers of this study acknowledge the important opportunities of the situation; potentially being able to identify and fill gaps of previous literature, while describing new openings for further research in the future.

Even though a significant difference was found between the text conditions and shadow and shadow/gaze respectively, it is important to remain critical towards the causality of the results. One critical note is that, not only did visual cues differ across conditions, a short text instruction was given along with the shadow and shadow/gaze condition as well, and this text instruction was not provided in the same manner as the text in the baseline condition. The text condition received a notable amount of clues compared to the other conditions, indicating that this condition is more difficult. In the baseline, participants received all information at once on how to solve the task, whereas the shadow and shadow/gaze conditions had ongoing pieces of text for each step towards solving the task following the cue. This difference might be decisive and cause a decreased isolation of the main effects.

Conclusion

The tutorial for Watchout was constructed to investigate the effect of visual cues in software learning, where it was hypothesised that shadowing and gaze following would reduce cognitive load and decrease time to task completion. Using an ANOVA, a statistical significance was found in time to task completion between the three conditions *text, shadow* and *shadow+gaze,* and a post hoc pairwise t-test identified a significant difference between *text* and *shadow* and *text* and *shadow+gaze,* however, not between *shadow* and *shadow+gaze.*

These results indicate that visual cues may have a positive effect on learning, however, due to study limitations this has to be validated further on a larger sample size. For future studies, it is suggested to isolate gaze following as a main effect, along with using a wider range of eye-tracking measurements.

References

Alemdag, E. & Cagiltay, K. (2018). A systematic review of eye tracking research on multimedia learning. *Computers & Education*, *125*. 413-428. https://doi.org/10.1016/j.compedu.2018.06.023.

Anderson, J. R. (2014). *Cognitive Psychology and Its Implications* (8. ed.). New York: Worth Publishers.

Canham, M., & Hegarty, M. (2010). Effects of knowledge and display design on comprehension of complex graphics. *Learning and instruction*, *20*(2). 155-166. https://doi.org/10.1016/j.learninstruc.2009.02.014.

Dataton. (n.d.). Dataton. Retrieved 12-05-2020 from: https://www.dataton.com/.

Figma. (n.d). 2020-06-03. https://www.figma.com/.

Glynn, S. M. & Di Vesta, F. F. (1979). Control of prose processing via instructional and typographical cues. *Journal of Educational Psychology*, *71*. 595-603. https://doi.org/10.1037/0022-0663.71.5.595.

Lorch, R. F., Lorch, E. P., & Klusewitz, M. A. (1995). Effects of typographical cues on reading and recall of text. *Contemporary Educational Psychology*, *20*(1). 51–64. https://doi-org.e.bibl.liu.se/10.1006/ceps.1995.1003.

Mayer, R. E. (2010). Applying the science of learning to medical education. *Medical Education*, *44*(6). 543–549. https://doi.org/10.1111/j.1365-2923.2010.03624.x.

Mayer, R. E., & Moreno, R. (2003). Nine Ways to Reduce Cognitive Load in Multimedia Learning. *Educational Psychologist*, *38*(1). 43–52. https://doi.org/10.1207/S15326985EP3801_6.

NASA. (2019). NASA TLX: TASK LOAD INDEX. 2020-05-21. https://humansystems.arc.nasa.gov/groups/TLX/.

Rickards, J. P., Fajen, B. R., Sullivan, J. F., & Gillespie, G. (1997). Signaling, note taking and field independence-dependence in text comprehension and recall. *Journal of Educational Psychology*, *89*(3). 508-517. https://doi.org/10.1037/0022-0663.89.3.508.

Schrepp, J. (1988). *User Experience Questionnaire Handbook,* (8th ed.). DOI: 10.13140/RG.2.1.2815.0245.

Van Gog, T., & Jarodzka, H. (2013). Eye tracking as a tool to study and enhance cognitive and metacognitive processes in computer-based learning environments. *International handbook of metacognition and learning technologies*. New York: Springer.

Watchout. (n.d.). Dataton. Retrieved 12-05-2020 from: https://www.dataton.com/products/watchout.