The effect of feedback in adaptive toolbars

BRIAN HANSEN and ANDERS GRØNNE
Aalborg University

Adaptive User Interfaces have been a subject of research since 1985. No existing research has attempted to draw the user's attention towards changes in the adaptive elements. An adaptive toolbar was implemented in a simulated text editor application. Feedback was used to inform the participant of updates in the adaptive toolbar. The effect of this feedback was tested using an eye-gaze tracker. It was found that feedback affected both the toolbar used by the participants to solve tasks, and in which toolbar the participants first sought the solution.

Categories and Subject Descriptors: D.D.2 [**Software/Software Engineering**]: Design Tools and Techniques; H.1.2 [**User/Machine Systems**]: Human Information Processing; H.5.2 [**User Interfaces**]: Graphical User Interfaces (GUI)

General Terms: Design, Experimentation, Human Factors

Additional Key Words and Phrases: Adaptive User Interfaces, Feedback, Eye-gaze Tracking

1. INTRODUCTION

The attention to usability in software development has steadily increased in recent years. As the complexity and functionality of software increases, so does the demands to the users if they are to use the software to its full extend. However, private user are not likely to follow courses or tutorials on all the applications they wish to use. One way of avoiding this training is to consider the user when designing the interaction. However, as humans are diverse by nature, it is not possible to design a single optimal interaction - one would potentially be required to design one for each user. This issue has been the cause of some frustration as illustrated by Benyon [1993]:

> *Attempting to selectively breed humans so that particular characteristics become dominant may be one way of improving usability, but one that would probably be deemed politically unacceptable.*

An alternative approach to increase the usability of an interaction would be to adapt relevant components of the graphical user interface based on its user. Research on these types of interfaces - know as Adaptive User Interfaces (AUIs) - began in 1985 with the work of Greenberg and Witten [Gajos et al. 2006]. Since then, the increase in computing power now enables algorithms to create a unique profile of the current user on the fly.

Research on AUIs has until now primarily concerned itself with the algorithms, the graphical design of AUIs and what affects user acceptance of AUIs. One aspect, which so far seems to have been disregarded by researchers, is transparency of interaction. Transparency of interaction was identified as an important part of Human-Computer Interaction (HCI) research in a report sponsored in part by the

Authors' address: School of Information and Communication Technology, Aalborg University, Fredrik Bajers Vej 7, 9220 Aalborg, Denmark.

US National Science Foundation. More explicitly the report stated that:

> "*The aim of HCI research should be to understand the principles behind what is necessary for transparent interaction.*" [Strong 1995]

Wensveen et al. [2004] identified feedback as an essential factor when creating transparency of interaction. None of the explored research on AUIs investigated the effect of providing feedback. This lack of research into the effect of using feedback in AUIs was the motivation behind this project. As such, the purpose of this project was to investigate if feedback could be linked with Adaptive User Interfaces in a meaningful way.

1.1 Previous work

The overall usefulness of AUIs remain the subject of some debate. An improved interaction is one advantage of AUIs, whereas the confusion caused by the inherent unpredictability of adaptive interfaces is a disadvantage [Gajos et al. 2008].

Greenberg and Witten [1985] were the first to introduce the notion of adaptive interfaces. They employed a user modelled, menu-driven selection to retrieve entries from a phone directory and showed an improvement when compared to a non-modelled menu. Benyon [1993] conducted a literature review where he identified individual differences in both cognition and personality, which influenced how individuals approached computer interaction. He further suggested how to cope with these differences in an AUI.

The effect of creating an adaptive component in textual menus was investigated by Findlater and McGrenere [2008]. Although the focus of their study was small monitors, a comparison of textual menus on large monitors (18" monitor) was conducted to have a baseline for evaluating the small monitors. The baseline study investigated performance of textual menus with an adaptive component, where the adaptive component had an accuracy of either 50 % or 78 % versus a textual menu without an adaptive component. They found no difference in efficiency (average time spent solving a task) between the non-adaptive and high accuracy conditions, whereas the low accuracy was significantly less efficient than the two other conditions.

How the amount of suggestions and how the accuracy of the suggestions in AUIs influenced the efficiency of the participants in textual menus was investigated by Tsandilas and Schraefel [2005]. It was found that as the accuracy of the predictions from the algorithm rose so did the efficiency of the participants. Also, as the amount of system suggestions was lowered the participants' efficiency increased.

Gajos et al. [2006] tested three different methods for implementing adaptive toolbars in altered versions of Microsoft Word 2003 and compared these to the original version. The adaptive methods were:

Moving: Items were moved from their original location to the adaptive toolbar.

Split: Items were copied from their original location to the adaptive toolbar.

Visual Popout: Items were highlighted with a different colour.

Participants were asked to rate the different interfaces based on a series of statements. A significant preference for the *Split Interface* was found when compared to the *Visual Popout Interface* and the original version. No significant difference was found between the *Split Interface* and the *Moving Interface*. A second experiment was conducted to obtain performance measurements. It was found that the *Split Interface* was more efficient than both the *Moving Interface* and the original version. The *Visual Popout Interface* was not included in the second experiment. Gajos et al. [2008] further elaborated their original study to investigate the effect of algorithm accuracy. They found that accuracy does have a significant impact on efficiency. This supported the findings by Findlater and McGrenere [2008]. It remains unclear exactly how the *Visual Popout* effect was implemented in the experiment, as this was not clarified in the article.

Wensveen et al. [2004] proposed a framework which can be used to evaluate interactive products. The framework described a model where feedback and feedforward were coupled through six different aspects:

Time: Action should follow function without delay.

Location: The function of the product should occur in the same location as the action.

Direction: The direction of the function should be identical to the direction of the action.

Dynamics: The dynamics (e.g. velocity) of the function should correspond to that of the action.

Modality: The modality, in which the function was provided, should be a logical consequence of the modality, in which the action was performed.

Expression: The expressive method, in which the action was given, should be reflected in the expression of the corresponding function.

Wensveen et al. [2004] argued for the importance of being mindful of these aspects when attempting to logically couple user actions with product functions. For example, in an adaptive toolbar the feedback regarding content updates should be presented at the time of the update, at the location of the toolbar, and in the same modality (visual). The framework also includes three forms of feedforward/feedback. These were:

Functional A consequence of the product performing its intended function, e.g. coffee dripping into the coffee pot.

Augmented Additional information provided by the product besides its intended function, e.g. the “On” LED lighting up after turning the coffee machine on.

Inherent The information provided as a natural consequence of performing the action, e.g. the motion of the button of the coffee machine when pressed.

The framework in its entirety can be seen in Figure 1. For instance, a pair of scissors has a direct connection between action and function. Whereas a purely electronic product may have no direct connections between action and function. This would represent the worst possible interaction conceivable.

2. METHOD AND MATERIALS

Several studies have investigated the usability aspects of an adaptive component in adaptive user interfaces [Gajos et al. 2006; Gajos et al. 2008; Tsandilas and

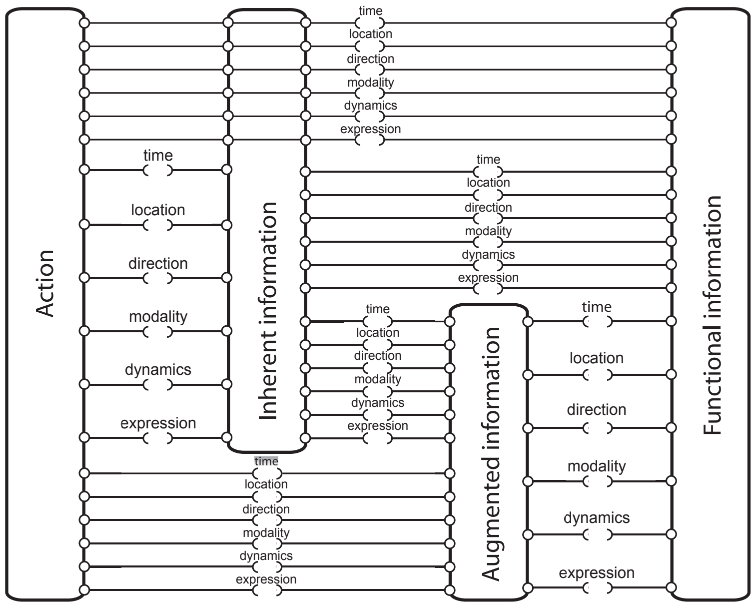

Fig. 1. The Interaction Frogger framework.

Schraefel 2005; Findlater and McGrenere 2008]. None of these studies attempted to draw the attention of the user to the adaptive component when it was updated. It is assumed that if the user is made aware of changes in the adaptive toolbar they will make use of the toolbar more often compared to a non-feedback version. Hence, the hypothesis for the experiment was:

Providing feedback on changes in an adaptive toolbar will increase the use of the toolbar.

2.1 Method

The participant was asked to locate and click a specific button in a simulated text editor. The text editor had static toolbars in the top left hand part of the interface and an adaptive toolbar in the top right hand side of the interface. These were separated by a 20^o visual angle to avoid any accidental discovery of a button without actively shifting gaze. This was inspired by the method used by Gajos et al. [2008]. The sought button would be located in the adaptive toolbar in twenty of thirty tasks to simulate an algorithm with this accuracy level. The design was implemented as a *Split Interface*, as Gajos et al. [2006] showed a significant preference for the *Split Interface* compared to a non-adaptive version. A screenshot from the test application can be seen in Figure 2. An eye-gaze tracker was employed to determine in which area of the text editor (static or adaptive) the participant first sought the required button. Each participant tested two versions of the text editor. The feedback version attempted to draw the attention of the participant to the adaptive toolbar when this was updated. The non-feedback version made no attempt to draw the attention of the participant. The feedback was implemented as a pulsating blue glow around the adaptive toolbar. The glow would pulsate twice. The initial glow would start 0.5 seconds after the participant had clicked the *Start* button. One pulse had a duration of 0.9 seconds and the pause between the two pulses had a duration of 0.7 seconds. The delay prior to the pulsating glow was implemented as it is the onset of motion that attracts attention and not the motion itself [Abrams and Christ 2003]. This implementation adhered to the *Time*, *Location*, and *Modality* aspects of the Interaction Frogger framework.

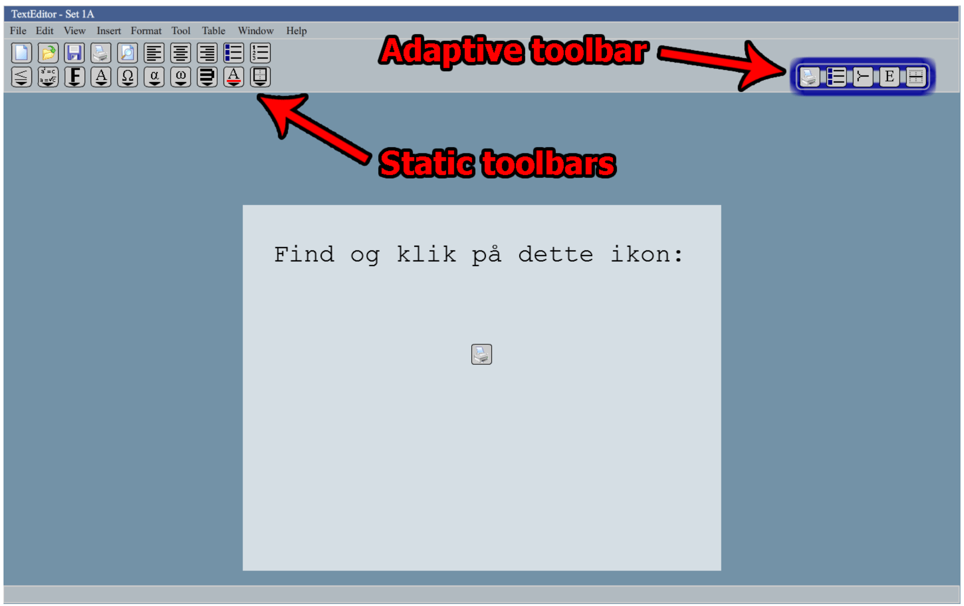

Fig. 2. The test application with the feedback present.

The experiment was a within-subject design. Before the actual experiment started the eye-gaze tracker was calibrated to each individual participant. After the calibration, the participant was required to read an introductory text, explaining their part in experiment. Half of the participants completed tasks for the feedback version of the interface first, the other half started with the non-feedback version. Two sets of fifteen tasks were created. This was done to prevent the participant from remembering the location of specific buttons. The specific tasks across task sets had an identical level of complexity, i.e. they did not vary in regards to whether the button was located directly on the interface or had to be accessed via a drop-down menu. If the button was in a drop-down menu, the menus would contain the same amount of buttons. For example, Task 2 in both task sets required the participant to find a button in a drop-down menu with 2x5 buttons. The order of feedback and task set were counterbalanced using a Latin square design. For each task the participant was shown a picture of a button, which they had to click - in either the static toolbars or the adaptive toolbar - to complete the task. The participant solved all tasks using a mouse as the only input device. This, to keep the participant's gaze on the screen. Also, this prevented the participant from trying to solve the tasks using keyboard shortcuts. After all fifteen tasks had been completed in one version of the text editor, the participant rated the 4 Likert statements seen in Figure 3. The statements translates from Danish to:

The adaptive toolbar was useful.

I knew beforehand when the required button was in the adaptive toolbar.

The adaptive toolbar made solving the tasks more efficient.

I felt in control of what occurred on the interface.

Upon completion of all 30 tasks the age and gender of the participant was logged. Also the participant rated their experience with text editors like Microsoft Word and OpenOffice.org on a five step scale ranging from novice to expert.

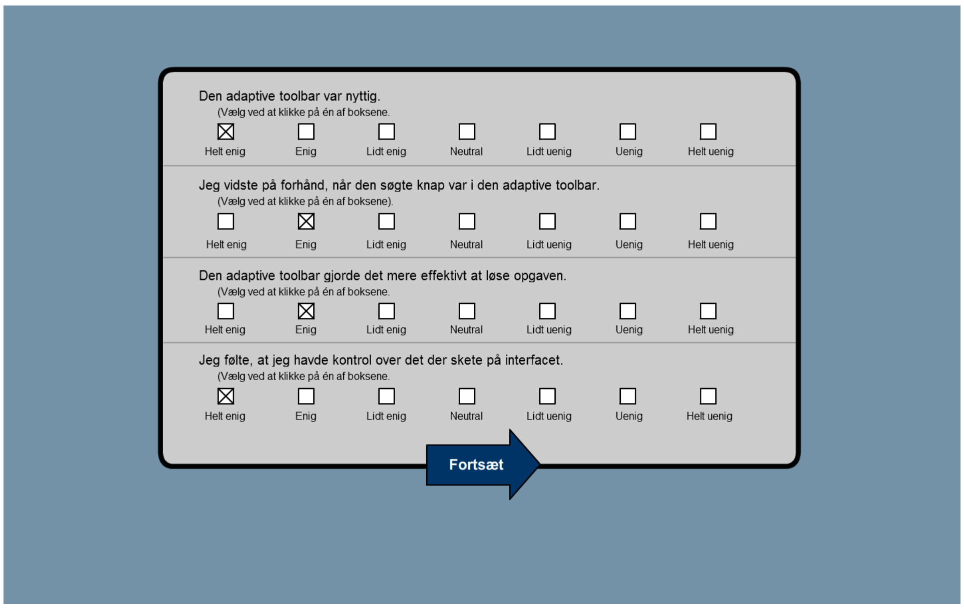

Fig. 3. The questionnaire presented to the participant after each task set had been completed.

2.2 Materials

A test application was developed in Adobe® Flash® Creative Suite, 4^{th} edition. The test application handled instructions and logging of user actions. A Tobii X120 eye-gaze tracker running Tobii Studio version 1.7.3 was used to record participant eye-gaze data.

3. RESULTS

The data from the experiment was analysed based on two measurements. These were: (1) the toolbar used to solve the task and (2) the toolbar which the participant fixated on first. These measurements were analysed in three steps. Firstly, linear regression was performed to investigate which variables influenced the above mentioned measurements. Secondly, model reduction was performed to eliminate uncertainties in border regions. Thirdly, the influencing variables were evaluated to determine how they influenced the participants' decisions. The variables, which were considered to possibly influence the participants' decisions, were:

Menu depth: Whether the required button was located in a drop-down menu or in the top level of the interface.

Feedback: Whether or not feedback was provided during the specific task.

Task: The task number.

Task set: The task set which the current task was in. Each task set contained 15 tasks.

Experience: The participant's self evaluated experience with text editors.

Age: The participant's age.

Gender: The participant's gender.

ID: A unique identifier for each participant.

All statistical tests in this analysis rejected the null-hypothesis at the 0.050 significance level.

3.1 Participants

All participants were recruited at the School of Information and Communication Technology at Aalborg University. All but one were students. 20 participants (15 male, 5 female) participated in the experiment. The age span was 20 to 33 years ($median = 24$, $sd = 0.55$). On the five step scale, two participants rated themselves as having a text editor experience level of three, fourteen participants rated themselves as four, and four participants rated themselves as having an experience level of five.

3.2 Toolbar usage

For each task the toolbar, which the participant used to solve the task, was recorded. As the adaptive toolbar only contained the requested button in twenty of thirty tasks not all tasks could be solved using the adaptive toolbar. The tasks, which could not be solved with the adaptive toolbar, were omitted from this analysis. A generalised linear model was defined using the variables listed above. Model reduction was performed to identify the variables that influenced the participant's choice. The influencing variables are listed in Table I.

	Estimate	**Standard error**	**p-value**	
Feedback	1.135	0.023	0.002	**
Menu depth	2.119	0.274	< 0.001	***
Task	−0.065	0.030	0.028	*
Task set	1.890	0.563	< 0.001	***
ID	0.070	0.023	0.002	**

Table I. The variables that influenced which toolbar the participants used to solve the tasks.

To investigate how feedback affected which toolbar the participant used to solve the task, the data was split into two groups based on whether feedback was provided or not. The frequency of how often the participants used the adaptive toolbar was calculated for both feedback conditions. The same analysis was conducted for the menu depth. These proportions can be seen in Figure 4. As seen the frequency of using the adaptive toolbar was higher when feedback was provided compared to when it was not ($X^2 = 13.337$, $p < 0.001$). The same was observed when the button was located in a drop-down menu ($X^2 = 81.010$, $p < 0.001$). To investigate if feedback affected any possible development in toolbar choice over time, the data was split into four groups. These were:

Group 1: First ten tasks with feedback

Group 2: Last ten tasks without feedback

Group 3: First ten tasks without feedback

Group 4: Last ten tasks with feedback

The difference in proportions was calculated to investigate whether the participants changed behaviour in regards to which toolbar they used to solve the tasks when feedback was either removed or added depending on their starting condition. As seen in Table II no difference was found between Group 1 and Group 2. However, between Group 3 and 4 a significantly higher proportion of tasks were solved using the adaptive toolbar, as also seen in the Table II.

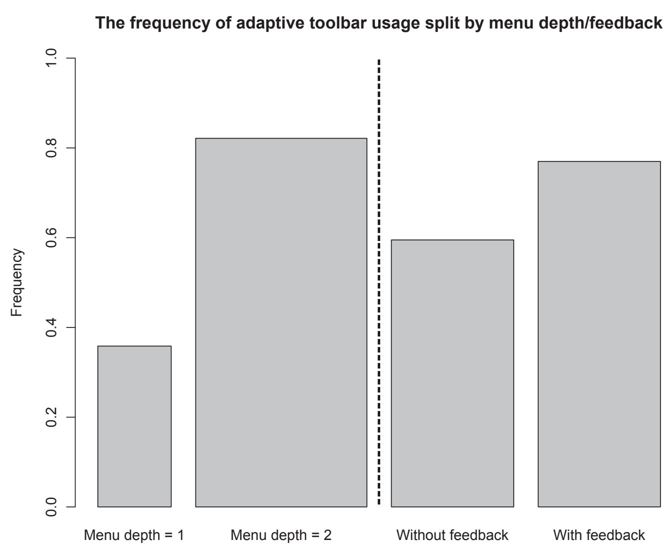

Fig. 4. The two coloums on the left show the difference in the proportions of tasks, which were solved using the adaptive toolbar, split by menu depth. The two coloums on the right show these same proportions, only split by feedback instead. The width of the bars is based on the number of trials for each condition.

	X^2	df	p-value		Alternative
Group 1 & 2 (Starting with feedback)	1.929	1	0.083		Group 1 > Group 2
Group 3 & 4 (Starting without feedback)	12.928	1	< 0.001	***	Group 4 > Group 3

Table II. The difference in proportions of usage of the adaptive toolbar between the four groups.

3.3 First fixation

For each task the eye-gaze tracker attempted to record where the participants first sought the requested button. In 25 out of a total of 600 tasks the eye-gaze tracker was unable to do so, and these 25 tasks were omitted from this analysis. A generalised linear model was defined using the variables listed in Section 3. Model reduction was performed to identify the variables that influenced where the participant's first sought the requested button. The influencing variables are listed in Table III.

	Estimate	Standard error	p-value	
Feedback	0.380	0.179	0.034	*
Task	0.049	0.011	< 0.001	***
Task set	0.756	0.179	< 0.001	***

Table III. The variables that influenced which toolbar participants gazed upon first.

To investigate how feedback affected where the participant first sought the requested button, the data was split into two groups based on whether feedback was provided or not. The result can be seen in Figure 5. A test of proportions revealed that the frequency of tasks, where the participant first sought the requested button in the adaptive toolbar, was higher when feedback was provided ($X^2 = 3.532$, $p = 0.030$).

To investigate if feedback affected any possible development over time the data was split into four groups. These were:

Group 1: First fifteen tasks with feedback
Group 2: Last fifteen tasks without feedback
Group 3: First fifteen tasks without feedback
Group 4: Last fifteen tasks with feedback

The results can be seen in Figure 6. The boxplots indicates that participants sought the requested button first in the adaptive toolbar more frequently in the second half of the experiment. This tendency was found significant, as seen in Table IV.

	X^2	df	p-value		Alternative
Group 1 & 2 (Starting with feedback)	5.372	1	0.010	*	Group 2 > Group 1
Group 3 & 4 (Starting without feedback)	24.324	1	< 0.001	***	Group 4 > Group 3

Table IV. The difference in proportions of those who first sought the requested button in the adaptive toolbar based on feedback condition and feedback order.

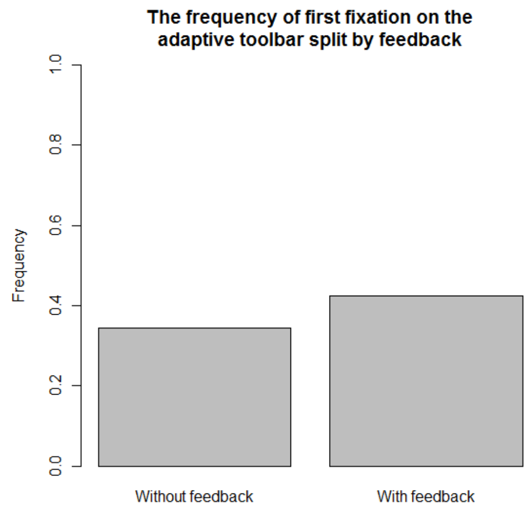

Fig. 5. Frequency of first fixations on the adaptive toolbar split by feedback.

The toolbar, which the participant had just used to solve the previous task, was compared to the one they would first seek the requested button in to examine any possible correlation. Task 1 and 16 were removed prior to this analysis. Task 1 was removed as no previous task existed which it could be compared to. Task 16 was removed since the participants had to answer a questionnaire after Task 15. The analysis showed that in 70.1 % of the trials, the toolbar, which the participant used to solve the previous task, was also the one the participant fixated on first. This was significantly above random ($X^2 = 77.864$, $p < 0.001$).

3.4 Likert analysis

The ratings of the Likert statements were investigate to examine whether feedback had an impact on user satisfaction rating. The participants' ratings to these questions are summarised in Table V. As seen, no clear difference exist between the ratings. That is, feedback did not seem to affect user satisfaction.

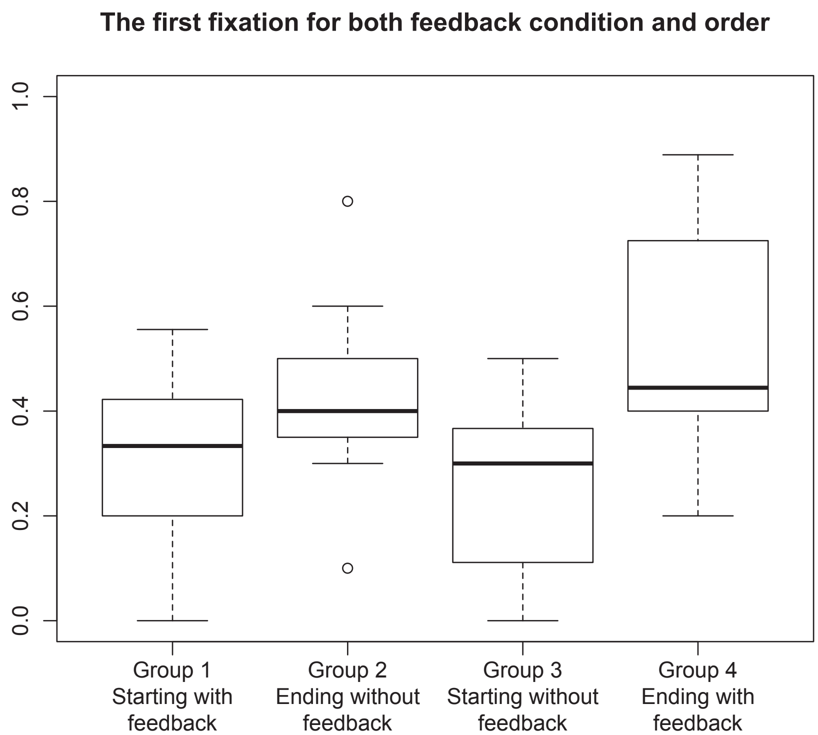

Fig. 6. Frequency of first fixations on the adaptive toolbar split by feedback and feedback order.

	Mean	**Standard deviation**	**Median**
Statement 1 with feedback	3.250	1.773	3
Statement 1 without feedback	3.450	1.099	3
Statement 2 with feedback	4.100	1.714	4.5
Statement 2 without feedback	4.350	1.531	4.5
Statement 3 with feedback	3.050	1.538	2.5
Statement 3 without feedback	3.600	1.188	4
Statement 4 with feedback	3.350	1.599	4
Statement 4 without feedback	3.150	1.461	3

Table V. The ratings to the questions asked in the experiment, grouped according to the feedback condition.

4. DISCUSSION

The main hypothesis *"Providing feedback on changes in an adaptive toolbar will increase the use of the toolbar"* is accepted. This acceptance is based on the correlation between feedback and both the use of the adaptive toolbar and the participants' first fixation.

There were no significant difference between the proportions of tasks solved using the adaptive toolbar when feedback was provided in the first fifteen tasks, compared to the last fifteen tasks where no feedback was present. However, the proportion of tasks where the participants, who used the adaptive toolbar to solve the tasks, was significantly higher for the feedback condition after having completed the first fifteen tasks without feedback. This indicates a clear correlation between usage of the adaptive toolbar and feedback. Also, since there is no detectable difference between the two tasks sets when participants started with the feedback condition, it indicates that the participants became aware of the toolbar almost immediately and continue to use it even after the feedback was no longer present. The second reason for accepting the hypothesis is that feedback was shown to have a significant influence on which toolbar the participants fixated on first. This effect is particularly apparent when reviewing the difference between Group 3 and 4 (right half of Figure 6). Furthermore, the participants who were provided with feedback initially continued to use the adaptive toolbar, even after the feedback had been removed. It can thus be concluded, that feedback does increase the use of an adaptive toolbar.

Another observation made in the toolbar usage analysis was the influence of menu depth. Participants were more prone to use the adaptive toolbar, when the button they sought was located in a drop-down menu. This is likely caused by the fact, that the top level of the static toolbar contains a certain range of standard text editor buttons such as *Save document*, *Print*, and *New document*. As the participants rated themselves as being of a moderate or higher experience level in typical text editors, their experience with these text editors will likely have provided them with a strong expectation of where these buttons were likely located. As such, it would be prudent to limit one's adaptive toolbar to only include buttons of a certain "complexity" in regards to menu depth. To put it another way, if participants have strong expectations of where a certain button is located, they will search for it there first regardless of feedback. Had the graphical content of all buttons been created to look nothing like any existing buttons, the usage of the adaptive toolbar may have been significantly increased. This finding supports one of the conclusions by Gajos et al. [2006]:

> *"We also demonstrated that the frequency of interaction with the interface and the cognitive complexity of the task influence what aspects of the adaptive interface users find relevant."*

A correlation between the toolbar, which the participant had just used to solve a task, and the toolbar, which they gazed on first in the following task, was observed. This occurred in a total of 70.1 % of all tasks. This tendency indicates that participants quickly developed expectations based on their previous success. As such, this underlines the necessity for high accuracy in the adaptive algorithms. If users find the adaptive toolbar lacking in terms of accuracy, it can quickly become more efficient for them to ignore the adaptive toolbar completely and simply search the static toolbar methodically.

The Likert ratings did not indicate any difference in the acceptance ratings in regards to the feedback condition. As such, the participants did not seem to be annoyed by the presence of the implemented feedback nor did it seem to have a positive effect on the participants' satisfaction rating of the adaptive toolbar. This is possibly influenced by the limited exposure. The long term effects of feedback would have to be explored to conclude whether users would eventually find the feedback a nuisance, or if the feedback was subtle enough to remain present. Either way, the conducted experiment indicates that only a limited use of feedback would be sufficient to heighten the use of an adaptive toolbar significantly.

An observation made during the review of the eye-gaze tracking recordings, was that participants would move their gaze towards either the static or adaptive toolbar immediately upon pressing the "Start" button. An example of this can be seen in Figure 7. This was curious as a small delay was present in the test application upon pressing this button. As a result hereof, participants would often stare at a black area on the screen for a few frames (33.33 milliseconds each frame) before the screen would change. This indicates that the participants had already decided where to search for the requested function button, before either of the toolbars were visible. As such, the participants would not have been exposed to the feedback as this occurred 0.5 seconds after the screen changed. If this phenomenon was present in more tasks than not, it could increase the necessity for feedback even further. As feedback can draw the gaze of users, it may draw their attention to a toolbar, which they had not noticed prior to the feedback appearing.

As the experiment was designed to test a specific hypothesis, it had to abide by certain design consideration. For instance, a large visual angle between the static and the adaptive toolbar was required to prevent participants from accidentally discovering the sought button. In an ecological experiment setting, the distance between the two toolbars would likely be smaller. This could potentially influence the effect of feedback - for good or bad - e.g. by forcing a user's gaze upon itself. It would also be prudent to adjust the algorithm to not include certain conventionally placed buttons, such as *Save document*, *Print*, and *New document*. Thus, limiting the content of the adaptive toolbar to only contain buttons with a certain complexity in regards to menu depth. An alternative use of the feedback effect could be to limit it to cases, where the adaptive toolbar has updated itself with high certainty of containing the button required by the user. As some predictions are bound to be more accurate than others, this could at the very least help improve the user acceptance of the adaptive toolbar.

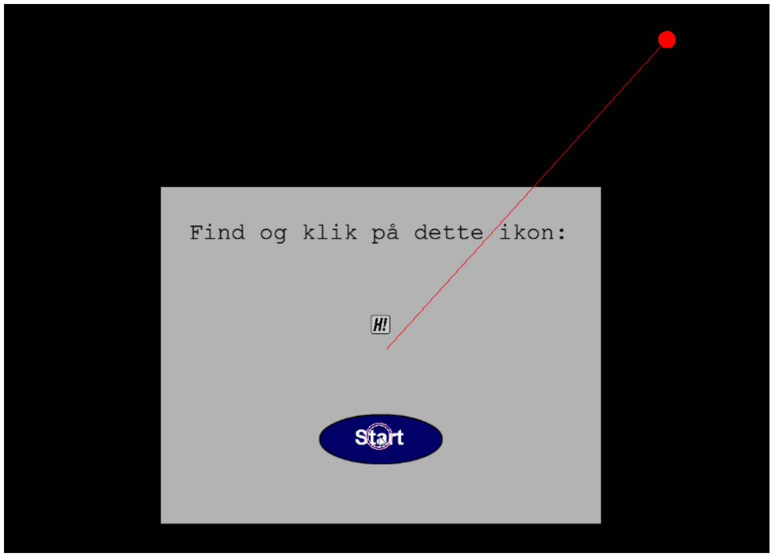

Fig. 7. A screenshot of a participant who looked at the adaptive toolbar after clicking the "Start" button, but before the screen changed.

5. ACKNOWLEDGEMENTS

The authors would like thank all those who participated in the experiment and those who assisted in the pilot studies of the experiment.

REFERENCES

ABRAMS, R. AND CHRIST, S. 2003. Motion onset captures attention. *Psychological Science 14,* 5, 427.

BENYON, D. 1993. Accommodating individual differences through an adaptive user interface. *HUMAN FACTORS IN INFORMATION TECHNOLOGY 10*, 149–149.

FINDLATER, L. AND MCGRENERE, J. 2008. Impact of screen size on performance, awareness, and user satisfaction with adaptive graphical user interfaces. In *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems.* ACM, 1247–1256.

GAJOS, K., CZERWINSKI, M., TAN, D., AND WELD, D. 2006. Exploring the design space for adaptive graphical user interfaces. In *Proceedings of the working conference on Advanced visual interfaces.* ACM, 201–208.

GAJOS, K., EVERITT, K., TAN, D., CZERWINSKI, M., AND WELD, D. 2008. Predictability and accuracy in adaptive user interfaces. In *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems.* ACM, 1271–1274.

GREENBERG, S. AND WITTEN, I. 1985. Adaptive personalized interfaces A question of viability. *Behaviour & Information Technology 4,* 1, 31–45.

STRONG, G. 1995. New directions in human-computer interaction: education, research, and practice. *interactions 2,* 1, 69–81.

TSANDILAS, T. AND SCHRAEFEL, M. 2005. An empirical assessment of adaptation techniques. In *CHI'05 extended abstracts on Human factors in computing systems.* ACM, 2009–2012.

WENSVEEN, S., DJAJADININGRAT, J., AND OVERBEEKE, C. 2004. Interaction frogger: a design framework to couple action and function through feedback and feedforward. In *Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques.* ACM, 177–184.