

---

# The Importance of Visual Attention for Adaptive Interfaces

**Fabian Göbel**

ETH Zurich  
Institute of Cartography and  
Geoinformation  
CH-8093 Zurich, Switzerland  
goebelf@ethz.ch

**Ioannis Giannopoulos**

ETH Zurich  
Institute of Cartography and  
Geoinformation  
CH-8093 Zurich, Switzerland  
igiannopoulos@ethz.ch

**Martin Raubal**

ETH Zurich  
Institute of Cartography and  
Geoinformation  
CH-8093 Zurich, Switzerland  
mraubal@ethz.ch

---

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.  
Copyright is held by the owner/author(s).  
*MobileHCI'16 Adjunct*, September 06-09, 2016, Florence, Italy  
Copyright © 2016 ACM 978-1-4503-4413-5/16/09.  
<http://dx.doi.org/10.1145/2957265.2962659>

**Abstract**

Efficient user interfaces help their users to accomplish their tasks by adapting to their current needs. The processes involved before and during interface adaptation are complex and crucial for the success and acceptance of a user interface. In this work we identify these processes and propose a framework that demonstrates the benefits that can be gained by utilizing the user's visual attention in the context of adaptive cartographic maps.

**Author Keywords**

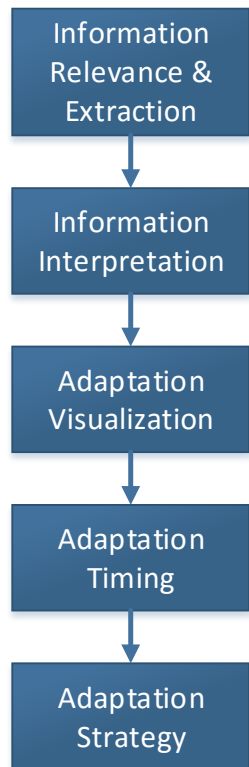
Gaze-Based Interaction; Visual Attention; Interface Adaptation.

**ACM Classification Keywords**

H.5.2 [Information interfaces and presentation]: User Interfaces

**Introduction**

User interaction with (mobile) computer systems is mostly facilitated through a graphical user interface. These interfaces often utilize an adaptation strategy in order to support the user in fulfilling her goals by adapting to her needs and current task. For instance, when a user is interacting with a cartographic map trying to find a nearby restaurant, the map interface might hide irrelevant information such as pharmacies or grocery stores.



**Figure 1:** Interface adaptation processes.

The processes that take place before an interface adapts are crucial for the success of an adaptation, i.e., how well the adjustment supports a user to achieve her goals. According to Hutchins et al. [5], an effective interface is one where the user has to interact only with task-relevant objects, therefore minimizing the cognitive load. Thus, the process of identifying and efficiently employing relevant objects is crucial for adaptive systems. Several sources and methods can be utilized to extract relevant information, such as interaction sequences with the interface, among others.

Today's progress in the area of eye tracking allows for a further source that can be efficiently integrated into the processes of adaptive interfaces. Eye tracking technology allows to extract the user's visual attention which can provide important information while interacting with the interface and the environment. For example, Giannopoulos et al. [4] tracked the user's gaze while interacting with a mobile map in order to adapt it based on the visual attention spent on specific map objects. Visual attention is often linked with the object being processed [7], faster than other pointing modalities as we usually look at a target first before interacting [12] and it is always "on" [6].

In this work we first identify the processes of adaptive interfaces (see Figure 1). Second, while prior work (e.g., Kiefer and Giannopoulos [9]) mainly focuses on the integration of visual attention into one specific aspect of the adaptation processes, we highlight the importance and give directions on how visual attention can contribute to each level of the processes. We introduce a framework that demonstrates in detail how each stage of deriving adequate adaptations can benefit from gaze data with the example of cartographic map interfaces.

## Application Scenario

Alice is interacting with a mobile map searching for an Italian restaurant. While she is looking at the map, our system detects her intention from her eye movements and the map information she spent attention on. Alice is having problems finding a restaurant that also meets further criteria (e.g., located close to a bus stop). After a short time period, the system recognizes that she is cognitively overloaded (e.g., based on her blinking frequency) and decides to provide her with an option for adaptation. Alice uses this option and the system starts the adaptation presenting the computed optimal visualization for her intention. With this help, Alice is able to easily find what she was looking for and happily walks towards this place.

## Processes Towards Gaze-Based Adaptation

The challenging part of adapting an interface is not only the choice of visualization type (e.g., altering the presentation), but also the underlying processes that take part, from identifying the relevant pieces of information to their interpretation. We define and discuss the five processes that are relevant for such a system in order to be able to determine the intentions of its user and adapt accordingly.

### *Information relevance and extraction*

The first process defines the relevant information space, i.e., the set of information objects that should be used by the further processes in order to extract knowledge. In the context of mobile maps, for example the buildings and streets displayed, might be relevant information [15]. Furthermore, details about the map extent could be also considered as relevant, e.g., the displayed country as well as whether it is an urban or rural area.

Next to the set of stimulus-based information, also the relevant user-based data has to be defined [16]. For instance,

the physical location of the user could also be considered relevant. Since we focus on gaze-based interface adaptation, the points of interest gazed at in both, the physical as well as digital environment may carry important context information.

*Information interpretation*

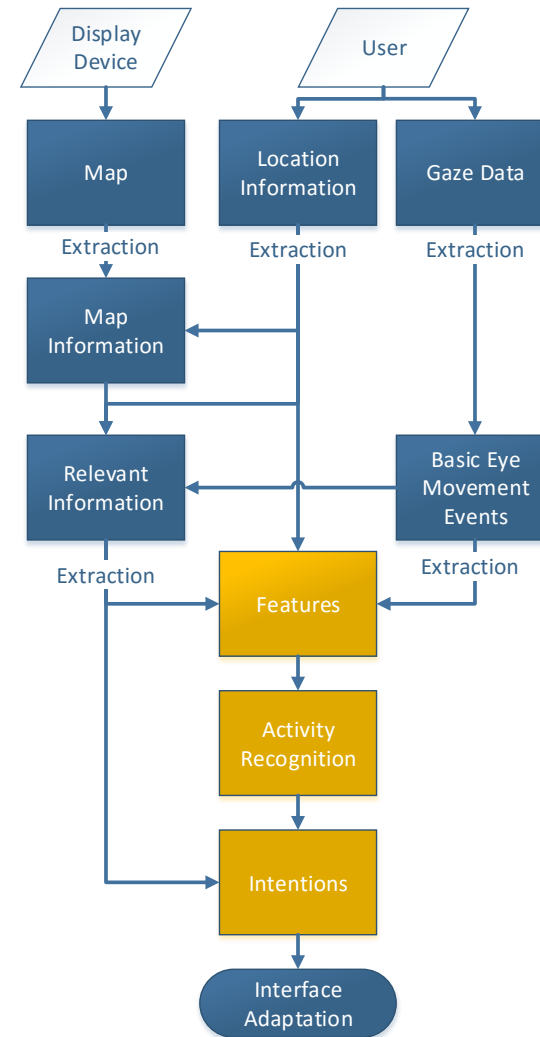
The extracted information can be interpreted through machine learning approaches. For instance, through content-independent approaches, e.g., gaze fixations, the activity of a user can be recognized. Furthermore, utilizing the relevant content, the user's intentions can be revealed that will allow us to prepare an optimal adaptation.

*Adaptation visualization*

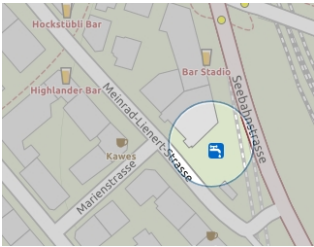
Even if we know what is relevant for the user as well as her intention, the adaptation visualization is crucial, since it has to be made in a way to help the user reach her goals, i.e., the adaptation must be in line with the user's mental model in order to optimally support her without interfering with her visual attention.

*Adaptation timing*

Although an adaptation is meant to help the users reach their goals more efficiently and effectively, if not well timed it might be confusing or even unwanted, since the users might feel patronized or out of control. As an example, consider a user trying to solve a task and while doing so, the interface changes. The objects of the interface a user might use as a reference frame would disappear, disorienting and probably frustrating the user. An adaptation would optimally occur when the user is cognitively overloaded but without interrupting the user's workflow. The user's eye movements (e.g., micro-saccades, blinks) can help identify whether she is cognitively overloaded and her visual attention can be utilized to detect the appropriate moment for adaptation.



**Figure 2:** Gaze-informed adaptation framework.



**Figure 3:** A map with roadways, POIs and labels. Only the parts that are gazed at by the user might be relevant. (source: opencyclemap.org)

### *Adaptation strategy*

The last challenge in the process is to choose the adaptation strategy meaning implicit or explicit adaptation [17]. Even if we are able to figure out when and how to adapt, it is important to decide whether the adaptation process starts automatically or whether the user should be given a choice for adaptation. For instance, the users might dislike a fully automated adaptation and instead prefer to get an option for adaptations (e.g., a button) making them feel more in control. In this process considering visual attention can help to assure that visual changes only appear in the area the user is currently focusing on to prevent disorientation.

## **A Framework For Gaze-Informed Adaptation**

As mentioned earlier, carefully selecting the information space for adaptations from the vast amount of available sources is crucial. In our framework, we focus on intelligent adaptations for digital maps (see Figure 2). In this context, the most valuable features are the user's spatio-temporal patterns and her gaze behaviour. From this information space we extract features to train machine learning algorithms for detecting activities and intentions. In the following we will describe the different processing steps of our framework in more detail.

### *Visual Attention*

The user's visual attention can be a great source of information for adaptive interfaces. Her visual focus tends to be related to comprehension and interest for the objects currently inspected [8, 3]. Three main types of eye events can be derived from gaze data: fixations, saccades and blinks [2]. Within a fixation, the eyes remain relatively still and focus on the point of interest. During this time period a visual stimulus can be perceived. The rapid eye movements between fixations are called saccades. They direct the current

point of regard towards a new target. A blink occurs when the eye lid rapidly closes and opens again, hiding the eyes. These eye events can be effectively utilized to retrieve the user's visual attention and infer interest as well as for measuring cognitive load [10].

### *The Map*

A map extent usually contains different features that are traditionally organized in layers (see Figure 3). The most often used ones are roadways, points of interest (i.e. Hospitals, Restaurants, Transportation), natural features, terrain, but also contour lines, boundaries and labels. Especially the features a user spends visual attention on can reveal important information about the user and her task.

### *Time and Location*

User specific spatio-temporal information, such as her current location and time zone can help specify the user context. For example, a map extent should adapt based on the user's location and current time to present valuable information. Furthermore, the user's visual attention spent in the environment from her current location could be incorporated into the process of "*information relevance extraction*" in order to identify further relevant data that can optimize the adaptation even more.

### *Task and Activity Recognition*

Activity recognition is based on observing the user behaviour and the environmental conditions to recognize specific actions. Especially while reading maps, analysing gaze movements seems to be very promising to identify actions and the steps users take to reach their goals.

Typically activity recognition is addressed with classifier learning [1, 14]. Kiefer et al. [11] computed 229 features based on saccades, fixations and blinks while performing one of six different map tasks. With these features they

trained a support vector machine. Although this approach does not consider the map features but solely relies on eye movement events, they could achieve a recognition accuracy of 77.8%. Extending this method by also incorporating the features of a map, that have been visually perceived, should increase recognition accuracy even further.

#### *Intention Recognition*

Whereas activity recognition aims to differentiate between several general tasks (e.g. focused search, route planning and polygon comparison), it is often not sufficient to support the user adequately. By only knowing the activity a user performs, we do not necessarily know which assistance is helpful. Intention recognition is an interpretation of the user's goals and mental state at a higher level. Often intentions and plans imply many different activities. Thus, inferring the intention requires not only to know activities but also their goals.

The map features a person focuses on can provide more precise information of the user's activity that is needed to, for example, not only allow to recognize that Alice is following a route, but also reveal she started from her home and finishes at a Italian restaurant in town. This knowledge is crucial for determining her intention and choosing the optimal adaptation.

#### *Interface Adaptation*

Interface adaptations need to meet the users expectations to be of help and ideally support the process of decision-making. To allow familiarizing with the interface and the environment, visual changes should occur only when the user is ready for them to take place. Furthermore, in the context of digital maps it is crucial to prevent over-simplification. For example, salient landmarks (e.g., churches) can significantly contribute to orientation. Eliminating them could counteract the idea of supporting the user's intention.

Considering gaze behaviour can be beneficial to address these challenges: On the one hand, features that are relevant for orienting will be looked at more often. On the other hand, eye events such as blinks are a sign of cognitive load. If they become less frequent, adaptation might be helpful.

Furthermore, gaze can be used to provide adaptations that might re-establish the context of reference [13] or reduce the search space [4] to find the position in case of an interruption.

### **Discussion and Conclusion**

In this paper we highlighted how tracking the user's gaze can contribute to answer the questions of how, when and where to adapt a user interface to allow for a more efficient and effective use.

We demonstrated the potential of considering the visual attention in each step of adaptation to create meaningful results. This was exemplified through a framework for information dense interfaces using the example of maps. Especially situations that require fast decision-making such as managing a crisis, public displays in train stations or interfaces with limited space like mobile phones can benefit from adequate interface adaptations.

Although this approach looks very promising, some limitations have to be addressed in future work: How to filter involuntary gaze movements? How to reliably identify objects inspected? What are convenient features to train machine learning algorithms? Do users accept these adaptations? How to achieve multi-user systems?

### **Acknowledgements**

This research has been supported by the Swiss National Science Foundation (grant number 200021\_162886).

## REFERENCES

1. A. Bulling, J. A. Ward, H. Gellersen, and G. Tröster. 2011. Eye Movement Analysis for Activity Recognition. *IEEE transactions on pattern analysis and machine intelligence* 33, 4 (2011), 741–53.
2. A. T. Duchowski. 2007. *Eye Tracking Methodology: Theory and Practice* (2nd ed.). Springer, London.
3. G. Edwards. 1998. A tool for creating eye-aware applications that adapt to changes in user behaviors. *In Proc. of the Third International Conference on Assistive technologies (Assets '98)* (1998), 67–74.
4. I. Giannopoulos, P. Kiefer, and M. Raubal. 2012. GeoGazemarks: Providing Gaze History for the Orientation on Small Display Maps. *In Proc. of the 14th International Conference on Multimodal Interaction (ICMI '12)*. ACM, 165–172.
5. E. Hutchins, J. Hollan, and D. Norman. 1985. Direct Manipulation Interfaces. *Human-Computer Interaction* 1, 4 (1985), 311–338.
6. R. J. K. Jacob. 1993. Eye Movement-Based Human-Computer Interaction Techniques : Toward Non-Command Interfaces. *Advances in human-computer interaction* 4 (1993), 151–190.
7. M. A. Just and P. A. Carpenter. 1976. Eye fixations and cognitive processes. *Cognitive Psychology* 8, 4 (1976), 441–480.
8. M. A. Just and P. A. Carpenter. 1980. A theory of reading: from eye fixations to comprehension. *Psychological Review* 87, 4 (1980), 329–354.
9. P. Kiefer and I. Giannopoulos. 2015. A Framework for Attention-Based Implicit Interaction on Mobile Screens. *In Proc. of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct (MobileHCI '15)* (2015), 1088–1093.
10. P. Kiefer, I. Giannopoulos, A. T. Duchowski, and M. Raubal. 2016. Measuring Cognitive Load for Map Tasks through Pupil Diameter. *In In Proc. of the Ninth International Conference on Geographic Information Science (GIScience 2016)*. Springer. to appear.
11. P. Kiefer, I. Giannopoulos, and M. Raubal. 2013. Using Eye Movements to Recognize Activities on Cartographic Maps. *In Proc. of the 21st International Conference on Advances in Geographic Information Systems (SIGSPATIAL '13)* (2013), 498–501.
12. M. Land, N. Mennie, and J. Rusted. 1999. The roles of vision and eye movements in the control of activities of daily living. *Perception* 28, 11 (1999), 1311–1328.
13. A. Mariakakis, M. Goel, Md T. I. Aumi, S. N. Patel, and J. O. Wobbrock. 2015. SwitchBack: Using Focus and Saccade Tracking to Guide Users' Attention for Mobile Task Resumption. *In Proc. of the International Conference on Human Factors in Computing Systems (CHI '15)* 1 (2015), 2953–2962.
14. J. Pärkkä, M. Ermes, P. Korpipää, J. Mantyjärvi, J. Peltola, and I. Korhonen. 2006. Activity classification using realistic data from wearable sensors. *IEEE Transactions on Information Technology in Biomedicine* 10, 1 (Jan 2006), 119–128.
15. J. Raper. 2007. Geographic relevance. *Journal of Documentation* 63, 6 (2007), 836–852.
16. M. Raubal and I. Panov. 2009. A Formal Model for Mobile Map Adaptation. *Location Based Services and TeleCartography II* (2009), 11–34.
17. A. Schmidt. 2000. Implicit human computer interaction through context. *Personal Technologies* 4, 2 (2000), 191–199.