# A Framework for Attention-Based Implicit Interaction on Mobile Screens

**Peter Kiefer**
ETH Zurich
Institute of Cartography and
Geoinformation
CH-8093 Zurich, Switzerland
pekiefer@ethz.ch


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

## Abstract

We propose to keep track of the user's attention during in-
teraction with a small display screen, and use that atten-
tion history for later interface adaptation. A framework for
attention-based implicit interaction is described that con-
sists of attention measuring, mapping, logging, aggregation,
interpretation, and interface adaptation. The framework is
exemplified by an interaction method described in previous
work: the GeoGazemarks approach.

## Author Keywords

Gaze-Based Interaction; Visual Attention; Information Over-
load

## ACM Classification Keywords

H.5.2 [Information interfaces and presentation]: User Inter-
faces

## Motivation

Although wearable devices, such as smartwatches and
head-mounted displays [24], are on the rise, the smart
phone with its 'one device fits all' property is still the most
prevalent mobile device these days. The high-resolution
displays, high processing power, strong multimedia capabil-
ities, good battery capacities, and the integration of numer-
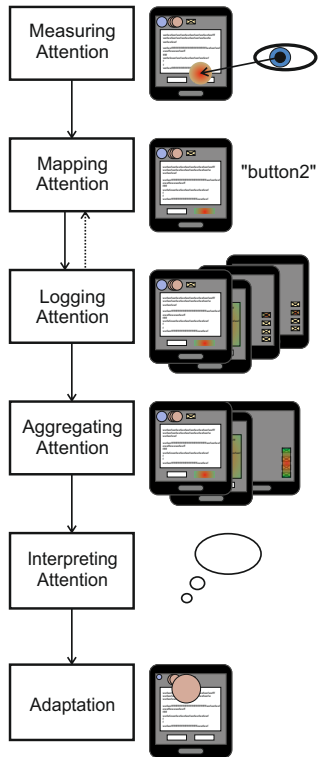ous sensors turn smart phones into everyday companions

**Figure 1:** Framework for Attention-Based Implicit Interaction on Mobile Screens

assisting with a plethora of mobile tasks, such as messaging, navigation, or gaming.

The diversity of tasks smart phones support, however, makes the interaction and interface design challenging. Complex, diverse, and frequently changing content needs to be accomodated on the very limited screen. An inadequate interaction and interface design may easily lead to information overload, even more since mobile tasks are often performed highly parallel and during interaction with the real world. Not surprisingly, cognitive load has been recognized as an important criterion for usability of mobile interaction [13].

Different strategies have been proposed for coping with information overload during the interaction with mobile devices, the most obvious one probably being the development of screen design guidelines for mobile devices [21]. Another strategy consists in adapting to the user's cognitive resource limitations based on context inference [2]. The Google Glass interface has explicitly featured short micro-interactions [24], while other wearable interfaces completely abandon the screen in favor for non-visual output, such as vibration [14, 11, 4]. Non-visual output modes, however, do not fit the general purpose requirement, and are hardly suited for content-rich interfaces, such as PDF reading, map exploration, or general web surfing.

In this paper, we propose to track the user's attention during interaction and use that attention history for later interface adaptation in order to avoid information overload. We describe a general framework for attention-based implicit interaction that manages attention with the following steps: attention measuring, mapping, logging, aggregation, interpretation, and interface adaptation. Note that, although most of this paper focusses on visual attention, we see at-

tention as a general concept in the line of attentive interfaces [25].

In the following section, we introduce the framework which is then exemplified by the GeoGazemarks approach in the Example section. We conclude the paper with an Outlook. Related Work is discussed throughout the paper whenever adequate.

## Framework

Figure 1 displays the processing steps of our framework which are explained in the following.

*Measuring Attention*
As a first step, the user's attention on the interface during normal interaction is measured. Different means can be used: attention data of high spatial and temporal resolution can be acquired by eye tracking, either through the front-facing smart phone camera [20], or through a head-mounted eye tracker [10, 22].

Eye detection is a more basic means of measuring visual attention, already implemented for current smart phones on the market[1]. Here, the presence of the eye in the front-facing camera is detected, but not the gaze position on the screen.

Next to measuring the visual attention directly, also non-visual interaction logs can serve as a proxy for visual attention. It is known that mouse and gaze positions show a high correlation in Desktop interaction [7]. Analogous, on a smart phone we can take the position of touch events as indication for visual attention. While eye tracking returns a stream of gaze points in screen coordinates (high resolution), the eye detection and touch events return (potentially
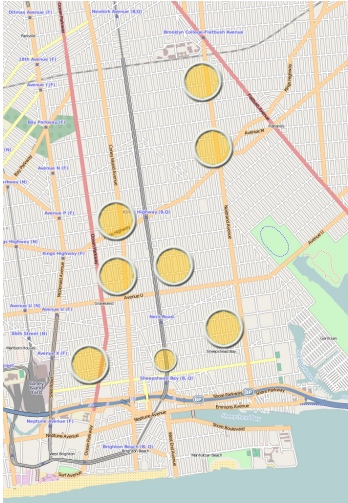
---

[1]e.g., the Samsung Galaxy S4

**Figure 2:** GeoGazemarks: Visual markers based on previous attention

fuzzy) regions on the screen (low resolution) as a basic unit for attention.

Note that the framework is flexible towards measuring attention on multiple screens [17]. The basic recorded unit of attention, such as a gaze position in pixel coordinates, would then be stored including information about the device it has been recorded on.

*Mapping Attention to Content*
Next, the screen coordinates must be mapped to the content that was visible on the screen at that moment. Depending on the resolution measured (point vs. fuzzy region), the density of the screen content, and the accuracy of the sensor, this mapping is not a trivial task. For cartographic maps, which contain specifically dense content, this problem has been investigated in [18] unter the title Gaze Map Matching. It has been argued that the temporal sequence of attention, combined with transition probabilites between content elements, can help for attention mapping. The dotted arrow in Fig. 1 indicates that using the log can help for the attention mapping

*Logging and Aggregation*
Having identified the content elements the user has attended to, these are logged to an attention history. Because the attention data come at high temporal resolution (e.g., eye tracking at 30 Hz), and because the memory of mobile devices is limited, a forgetting and/or aggregation of the attention history may become necessary. Deciding what to forget and/or aggregate depends on what we would like to do with the attention data later, which is described in the following section.

*Interpretation*
Attention can, first and foremost, be seen as an indicator for interest in a part of the interface. This interest is driven

by both, bottom-up and top-down processes [5]. The first are mainly determined by (visual) saliency while the second can be described as the user's activities, intentions or plans, depending on the level of abstraction. In order to provide an adequate adaptation it becomes necessary to determine which of these processes caused the visual attention, i.e., to interpret the user's attention log. This is an ongoing active research field in HCI [19, 6, 1, 3], and the algorithms used here are highly dependend on the interface type and tasks performed.

*Interface Adaptation*
By knowing which interface elements the user has paid attention to, and even more if we were further able to infer activities or intentions, intelligent interface adaptation can be designed. A number of papers in HCI investigate adaptive user interfaces (e.g., [16]). There is a clear agreement that adaptations must be designed carefully to avoid users feeling patronized, and to avoid causing information overload. Instead, the existing overload in the interface before the adaptation, such as, too many notification icons, should be reduced through the adaptation. The expected cost of the adaptation must be less than then the expected cost of not adapting [15]. Formal adaptation models are required which integrate a number of context parameters into the adaptation [23]. Moreover, adaptation will most likely change the user's attention, which will again be measured in the first level of the framework.

## Example
GeoGazemarks is a gaze-based implicit interaction concept that provides the user's visual attention history on a mobile map as visual clue to facilitate orientation [10]. The proposed framework captures and divides the complexity among the components relevant for the GeoGazemarks concept.

**Measuring Attention:** The visual attention in the GeoGazemarks concepts is recorded using an eye tracking device, returning gaze points in screen coordinates.

**Mapping Attention to Content:** These coordinates are transformed to geospatial coordinates (latitude, longitude) by mapping the visual attention to the geospatial content displayed at the recorded screen coordinates.

**Logging and Aggregation:** Next, the geospatial coordinates are aggregated to fixations and logged to the attention history. Fixations occur when the eyes remain relative still over a period of time.

**Interpretation:** A clustering algorithm was applied in order to aggregate fixations in spatial proximity. The size of the cluster is based on the amount and duration of the fixations. The more and longer fixations in the spatial proximity, the higher the interest for that coordinate. In GeoGazemarks, attention for a spatial coordinate on the map is interpreted as having interest for content at that coordinate. That means, it is assumed that visual attention includes perception, and that perception includes interest. Higher-level cognitive states, such as activities or intentions, are not the focus of GeoGazemarks.

**Interface Adaptation:** Finally, once the user of GeoGazemarks performs a zoom-out interaction, the map adapts, displaying the visual attention clusters on the map (see figure 2). This should help the user to orient based on the map areas she was previously spending her visual attention to.

## Outlook

As future work, the framework will be applied to interfaces in different domains, and different methods will be tested on the individual processing levels. Open questions to be investigated include: how reliant are the upper processing levels on a precise attention measuring sensor (or can we achieve good results also with other means of measuring attention)? What are good models and algorithms for interpreting attention in terms of activities/intentions? Here, we can rely on existing models from research on visual attention and cognitive models for the respective domain (e.g., [9] for airline pilots). How to design a good adaptation that is really perceived as helping, instead of annoying and patronizing?

While applying the framework to *multiple devices* is straightforward for the lower levels – attention measuring, mapping and logging – there is some challenge in the upper levels of the hierarchy: how can we aggregate and interpret attention during the interaction with multiple devices?

Extending the framework to attention-based interaction of *multiple users* is another interesting area for future research. One user could share her visual attention with others, e.g., for teaching purposes [8], or multiple users could share their attention to solve a task collaboratively [12].

## REFERENCES

1. Thomas Bader, Matthias Vogelgesang, and Edmund Klaus. 2009. Multimodal integration of natural gaze behavior for intention recognition during object manipulation. In *Proceedings of the 2009 international conference on Multimodal interfaces*. ACM, 199–206.

2. Jörg Baus, Antonio Krüger, and Wolfgang Wahlster. 2002. A resource-adaptive mobile navigation system. In *Proceedings of the 7th international conference on Intelligent user interfaces*. ACM, 15–22.

3. Roman Bednarik, Hana Vrzakova, and Michal Hradis. 2012. What do you want to do next: a novel approach for intent prediction in gaze-based interaction. In

*Proceedings of the symposium on eye tracking research and applications*. ACM, 83–90.

4. Mark Billinghurst and Thad Starner. 1999. Wearable devices: new ways to manage information. *Computer* 32, 1 (1999), 57–64.

5. Ali Borji, Dicky N Sihite, and Laurent Itti. 2014. What/where to look next? Modeling top-down visual attention in complex interactive environments. *Systems, Man, and Cybernetics: Systems, IEEE Transactions on* 44, 5 (2014), 523–538.

6. Andreas Bulling, Jamie A Ward, Hans Gellersen, and Gerhard Troster. 2011. Eye movement analysis for activity recognition using electrooculography. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 33, 4 (2011), 741–753.

7. Mon Chu Chen, John R Anderson, and Myeong Ho Sohn. 2001. What can a mouse cursor tell us more?: correlation of eye/mouse movements on web browsing. In *CHI'01 extended abstracts on Human factors in computing systems*. ACM, 281–282.

8. Shiwei Cheng, Zhiqiang Sun, Lingyun Sun, Kirsten Yee, and Anind K Dey. 2015. Gaze-Based Annotations for Reading Comprehension. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 1569–1572.

9. Florian Frische, Jan-Patrick Osterloh, and Andreas Lüdtke. 2011. Modelling and Validating Pilots' Visual Attention Allocation During the Interaction with an Advanced Flight Management System. In *Human Modelling in Assisted Transportation*. Springer, 165–172.

10. Ioannis Giannopoulos, Peter Kiefer, and Martin Raubal. 2012. GeoGazemarks: Providing Gaze History for the Orientation on Small Display Maps. In *Proceedings of the 14th International Conference on Multimodal Interaction (ICMI '12)*. ACM, New York, NY, USA, 165–172. DOI:http://dx.doi.org/10.1145/2388676.2388711

11. Ioannis Giannopoulos, Peter Kiefer, and Martin Raubal. 2015a. GazeNav: Gaze-Based Pedestrian Navigation. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM. to appear.

12. Ioannis Giannopoulos, Peter Kiefer, and Martin Raubal. 2015b. Watch What I Am Looking At! Eye Gaze and Head-Mounted Displays. In *Mobile Collocated Interactions, Workshop at CHI 2015*.

13. Rachel Harrison, Derek Flood, and David Duce. 2013. Usability of mobile applications: literature review and rationale for a new usability model. *Journal of Interaction Science* 1, 1 (2013), 1–16.

14. Wilko Heuten, Niels Henze, Susanne Boll, and Martin Pielot. 2008. Tactile wayfinder: a non-visual support system for wayfinding. In *Proceedings of the 5th Nordic conference on Human-computer interaction: building bridges*. ACM, 172–181.

15. Eric Horvitz, Andy Jacobs, and David Hovel. 1999. Attention-sensitive alerting. In *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 305–313.

16. Anthony Jameson. 2009. Adaptive interfaces and agents. *Human-Computer Interaction: Design Issues, Solutions, and Applications* 105 (2009).

17. Dagmar Kern, Paul Marshall, and Albrecht Schmidt. 2010. Gazemarks: gaze-based visual placeholders to ease attention switching. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2093–2102.

18. Peter Kiefer and Ioannis Giannopoulos. 2012. Gaze Map Matching: Mapping Eye Tracking Data to Geographic Vector Features. In *Proceedings of the 20th SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, New York, NY, USA, 359–368. DOI: http://dx.doi.org/10.1145/2424321.2424367

19. Peter Kiefer, Ioannis Giannopoulos, and Martin Raubal. 2013. Using Eye Movements to Recognize Activities on Cartographic Maps. In *Proceedings of the 21st SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, New York, NY, USA, 488–491. DOI: http://dx.doi.org/10.1145/2525314.2525467

20. Emiliano Miluzzo, Tianyu Wang, and Andrew T Campbell. 2010. EyePhone: activating mobile phones with your eyes. In *Proceedings of the second ACM SIGCOMM workshop on Networking, systems, and applications on mobile handhelds*. ACM, 15–20.

21. Jakob Nielsen and Raluca Budiu. 2013. Designing for the small screen. In *Mobile usability*. MITP-Verlags GmbH & Co. KG, Chapter 3.

22. Lucas Paletta, Helmut Neuschmied, Michael Schwarz, Gerald Lodron, Martin Pszeida, Patrick Luley, Stefan Ladstätter, Stephanie M Deutsch, Jan Bobeth, and Manfred Tscheligi. 2014. Attention in mobile interactions: gaze recovery for large scale studies. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*. ACM, 1717–1722.

23. Martin Raubal and Ilija Panov. 2009. A formal model for mobile map adaptation. In *Location Based Services and TeleCartography II*. Springer, 11–34.

24. Thad Starner. 2013. Project glass: An extension of the self. *Pervasive Computing, IEEE* 12, 2 (2013), 14–16.

25. Roel Vertegaal. 2002. Designing attentive interfaces. In *Proceedings of the 2002 symposium on Eye tracking research & applications*. ACM, 23–30.