

Preprint version. Springer, 2014. This is the authors version of the work. Please do not redistribute. The definitive version will be published in Multimedia Tools and Applications (DOI to be assigned).

Please cite as: Giannopoulos, I., Schöning, J., Krüger, A., & Raubal, M. (2014). Attention as an Input Modality for Post-WIMP Interfaces Using the viGaze Eye Tracking Framework. Multimedia Tools and Applications.

Attention as an Input Modality for Post-WIMP Interfaces Using the viGaze Eye Tracking Framework

Ioannis Giannopoulos, Johannes Schöning, Antonio Krüger, Martin Raubal

Abstract Eye tracking is one of the most prominent modalities to track user attention while interacting with computational devices. Today, most of the current eye tracking frameworks focus on tracking the users gaze during website browsing or while performing other tasks and interactions with a digital device. Most frameworks have in common that they do not exploit gaze as an input modality. In this paper we describe the realization of a framework named viGaze. Its main goal is to provide an easy to use framework to exploit the use of eye gaze as an input modality in various contexts. Therefore it provides features to explore explicit and implicit interactions in complex virtual environments by using the eye gaze of a user for various interactions. The viGaze framework is flexible and can be easily extended to incorporate other input modalities typically used in Post-WIMP interfaces such as gesture or foot input. In this paper we describe the key components of our viGaze framework and additionally describe a user study that was conducted to test the framework. The user study took place in a virtual retail environment, which provides a challenging pervasive environment and contains complex interactions that can be supported by gaze. The participants performed two gaze-based interactions with products on virtual shelves and started an interaction cycle between the products and an advertisement monitor placed on the shelf. We demonstrate how gaze can be used in Post-WIMP interfaces to steer the attention of users to certain components of the system. We conclude by discussing the advantages provided through the viGaze framework and highlighting the potentials of gaze-based interaction.

Keywords prototype multimedia systems and platforms · human-computer interaction · eye tracking

1 Introduction

With the increase of computational power and the quality of sensors that are available, pervasive systems that are embedded in our environment can provide us with various services to support our daily routines and activities. Therefore human attention becomes the critical resource in pervasive system design today [8]. Knowing whether a user is paying attention to a system or to which subparts of a system could significantly improve the usability and user experience of current pervasive systems.

Ioannis Giannopoulos & Martin Raubal
Institute of Cartography and Geoinformation, ETH Zurich, HIL, Wolfgang-Pauli-Strasse 15, 8093 Zurich, Switzerland
E-mail: igiannopoulos@ethz.ch, mraubal@ethz.ch

Johannes Schöning,
Hasselt University - tUL - iMinds, Expertise Centre for Digital Media, Wetenschapspark 2, 3590 Diepenbeek, Belgium
E-mail: Johannes.schoening@uhasselt.be

Antonio Krüger,
DFKI GmbH, Campus D3 2, Stuhlsatzenhausweg 3, D-66123 Saarbrücken, Germany E-mail: krueger@dfki.de

DRAFT

Users' gaze is a strong indicator for finding out whether users are paying attention to a system or not. Eye tracking has been employed to investigate how users interact with computational systems, e.g. what the results page of a search engine should look like. In contrast to research that focuses on where and how users are paying attention, we focus on how attention, in form of gaze fixations [6], can be used as an input modality. This resulted in the development of the viGaze framework. In viGaze the users gaze can be used as an input modality to steer interactions within a virtual pervasive environment. Our framework can be utilized for a wide range of pervasive applications and reduces the efforts of designing and developing various interaction patterns. We evaluated viGaze in a virtual pervasive retail environment as this provides a good example of a future pervasive environment cluttered with technology.

Imagine a future grocery store, which contains various sensors, displays and actuators to enrich and support the shopping experience as envisioned by Krüger et al. [15] (see figure 1). The users are surrounded by digital signage, electronic price tags and technology sensing their shopping behavior. The modeling of such complex virtual pervasive environments can be achieved using our framework, with low effort. We can also model complex interaction in such environments with relatively low effort; e.g. we can model how a users gaze can provide dynamic interactions, such as starting a video advertisement on a digital signage or a blinking pattern on one of the electronic price tags, in order to switch attention to the products of interest. This demonstrates how viGaze can enable novel ways of explicit and implicit interactions between customers and the retail environment. Our research contributions consist of:

- The design and implementation of the viGaze prototype framework, that focuses on providing an easy to use framework to exploit the use of eye gaze as an input modality (among other input modalities) in various contexts.
- A proof of concept of how viGaze enables explicit and implicit interactions within a pervasive retail environment.
- A study that shows the effectiveness of these interaction patterns.

2 Related Work

The related work section is subdivided into two main parts to allow a better overview on research areas where eye tracking is currently used within the field of psychology and human-computer interaction (HCI) for different purposes.

2.1 Visual Attention

In our daily environment, a massive amount of signals are provided to us through our eyes, but usually we process only the few ones that reach our awareness, and the ones that reach our memory [20,25]. The decision about what information entering our eyes gains access to our visual information processing system for further processing is taken by a selection mechanism. This selection process is controlled by cognitive factors as well as sensory stimulation and the dynamic interaction of these factors controls where and on what we pay attention [4]. Through this interaction, a saliency map of the visual environment is created, that highlights regions of interest as an aid to shift the gaze to these locations [25]. Visual attention can also be influenced by several factors such as sound, time and task constraints and age of the users [11,17]. In the user study reported in section 4, we used attention grabbing elements such as sound and visual effects in order to influence the visual attention of the users.

2.2 Eye Tracking in HCI research

Eye tracking devices can be used as input media in computer systems in order to allow easy, natural, and fast ways of interacting with virtual environments by using the gaze points of the users as input [12,24]. In

DRAFT

fact, there has been research addressing the abilities provided by an eye tracker for investigating several novel interaction techniques. An interface aware of its users gaze orientation can support user interaction with extended context based information such as translations of foreign words that are read in a text [1], with eye-driven cursors that behave differently according to context [18], or with assisting functionalities such as the placeholder functionality, called *GazeMarks*, that was proposed in [13]. The latter idea was also transferred to the spatial domain with the GeoGazemarks approach, where a users gaze history is utilized for more efficient orientation on small display maps [9]. In [1], the *Text 2.0* Framework was introduced, which is a browser plug-in that enables web pages with gaze handling functionalities allowing them to react to gaze. The *eyeBook*, which is an application that uses augmented text to respond with real time effects according to the current reading position, is an interesting example of the eye tracking interaction capabilities. Eye movements can also be used as gestures in order to enhance context aware gaming experience as shown by Bulling et al. [2]. Stellmach et al. [23] investigated the potential of gaze-supported panning in combination with different zooming modalities such as traditional mouse input or gesture based and touch interfaces. Duggan and Payne [7] investigated readers eye movements as they skimmed through expository text under time pressure. More generally, Komogortsev et al. [14] considered the effects of user interface adaptation based on regional eye tracker accuracy to improve user performance and satisfaction in an gaze-guided application. Cauchard et al. [3] used an eyetracker for measuring switches in visual context between displays in mobile Multi-Display Environments. Santella et al. [21] provided an interactive method for cropping photographs given minimal information about the location of important content, provided by eye tracking. Ohno et al. [16] presented a system that enables the user to interact with the computer after catching the users eye blinks.

In the mentioned research examples above, whenever gaze was used as an interaction modality to trigger an event, the system used had to be implemented from scratch. Most systems that enable gaze as an input modality are optimized for the investigation of the areas of interest (AOI), rather than for the investigation of interaction patterns. The environments that can be utilized are mostly limited to static images or videos, without using gaze to trigger events. Moreover, these systems are limited to gaze input, not allowing research on multimodal approaches. To our knowledge viGaze is the first framework that supports gaze as an input modality in dynamic virtual environments for interaction research purposes. This opens up new possibilities for other researchers in various domains.

3 viGaze Framework

In the following we describe the four main components of the viGaze eye tracking framework as shown in figure 2. A pervasive retail environment is used to illustrate the different components as examples. The framework can also be easily adapted to other interaction concepts. We picked the retail domain as it provides an interesting and rich testbed for our viGaze framework.

3.1 Virtual Environment

The first component, *Virtual Environment* (see figure 2, top) is used to easily model or integrate various virtual pervasive environments into the framework. In our scenario we modeled a grocery store shelf with an interactive display and electronic price tags (see figure 3) to enrich and support the shopping experience as described in [15,22]. In the first step, the background of the virtual environment had to be set. In our case we used an image of a real shelf. In the second step, through a simple drag and drop functionality we enhanced the shelf with different virtual products (image files) as well as several multimedia effects (flash files) on the price tag locations (see figure 4). In general, the virtual environment consists of a series of layers with images and flash objects. Our shelf can be enhanced with these two data types in order to simulate static objects like products (images) or dynamic objects like digital price tags and advertisement screens (flash). All the objects placed on the shelf can be freely moved in all directions (also in depth), allowing for customization. Next to the self designed environments, the framework allows the integration of existing environments. This integration process follows standard implementation and

DRAFT

binding routines according to the provided SDK. The Google Earth environment was easily integrated into the framework in order to evaluate multimodal interaction approaches with the virtual globe [10].

3.2 Input Media Library

The second component is the *Input Media Library* (see figure 2, right). It can be used to integrate a range of input devices such as an eye tracker or other Post-WIMP input devices. Although we mainly focus on eye tracking technology as an input and interaction modality, the *Input Media Library* can also integrate various input devices and techniques that are used in Post-WIMP interfaces. Currently the viGaze framework supports input using a mouse, foot input, the Wiimote and the Kinect besides gaze input. These input devices can be integrated through their SDKs and connected to an interface provided by the framework that handles the input events. This interface can be used for multimodal fusion and provides access to the virtual environment in order to trigger events. This opens up different ways for exploring multimodal approaches. The Wii Balance board and the X120 Tobii Eye Tracker were integrated through this library in order to evaluate multimodal interactions utilizing foot and gaze input [10] similar as proposed by Daiber et al. [5].

3.3 Interaction Library

Using the third component, called *Interaction Library* (see figure 2, left), the interactions and relationships between the various sensors and actuators in the virtual pervasive environments can be defined. This library provides a method where the logic of an interaction pattern can be implemented. For instance, if a gaze fixation serves as a trigger, this method will be called in order to activate a reaction of the environment. This component enables access to the data structures of the virtual environment, providing control over all objects (i.e., image and flash objects) that are used to define implicit and explicit interactions in the environment. Using this library any kind of interaction can be designed to guide or steer a user. One can imagine various possible interactions. Below we list two interactions for a virtual shelf in a pervasive retail environment. Both are evaluated within the user study described later in the paper (see section 4).

Example 1: If a gaze fixation occurs on a product in the shelf, the shelf triggers an advertisement on the display (see figure 5, left).

Example 2: When the user’s attention is shifted to the display for a period of time, the users gaze can be redirected to the product by letting the price tags under the product blink (see figure 5, right), or even highlight the path to an object by activating all the effects that are on the visual path to the object (see figure 6).

3.4 Analytics Tool

The final component, the *Analytics Tool* (see figure 2, bottom), can be used to easily run and evaluate user studies for testing the effectiveness of the interaction patterns defined in the *Interaction Library*. In order to be able to easily perform user studies, we created a user study engine that automates and coordinates experiments. This component can be used to register a user as well as to define the order in which a series of virtual environments is displayed. It logs all data generated during the experiments and provides a basic report and statistics functionality.

3.5 Application Scenario

In this section we provide an application scenario that highlights the functionality of the viGaze framework as well as the interplay between the four components.

DRAFT

In order to set up a user experiment, each of the four components has to be used. The first thing that is necessary in order to perform a user study is a virtual environment. Either a previously created shelf configuration will be loaded from the database or a new one has to be created. In order to create a new shelf configuration it is necessary to load the desired product images and drag them at the desired location on the shelf where they can also be freely rearranged. Afterwards, the flash files that simulate the digital price tags and the advertisement monitors have to be loaded and again, similarly to the product placement method above, they have to be placed on the shelf.

In a second step, the input device that is desired has to be selected through the main interface or integrated using the Input Media Library. This library provides an interface where the implementation of this device can be specified by using predefined methods. For instance, we can use the coordinates retrieved from the eye tracker and perform a click event whenever a gaze fixation occurs. If we would also use another device such as the Microsoft Kinect camera, we could combine the coordinates retrieved through both devices and enable a multimodal event.

In the third step, it is necessary to define the interaction patterns that are under investigation. The Interactions Library provides an interface where the interaction pattern can be designed. This Interface has access to the data structures of the virtual environment (i.e., products, digital price tags and advertisement monitors) and through its predefined methods, the logic of the interaction pattern can be specified. This interface is coupled with the interface provided by the Input Media Library and for instance, based on an event triggered by an input device, can enable and disable the effects on the shelf.

Finally, the last component can be used to set up a user study by selecting the configurations of the environment and interaction patterns to be tested as well as to define the duration and order of appearance of the shelves. Using the function keys of the keyboard, the experimenter can control the experiment.

3.6 Implementation

In this section the implementation details of the viGaze framework will be introduced. All four components of the described framework were implemented in order to have a prototype system to conduct user studies.

Initially, the viGaze framework was developed for experiments with the Tobii X60 eye tracker. In order to use the full capabilities of the Tobii SDK, the framework was developed in VB.NET. The libraries of the viGaze framework provide several methods through Interfaces in order to integrate, enable and control input media as well as to design interactions. Currently, in order to integrate a virtual environment such as Google Earth, or an input device such as the Microsoft Kinect camera, it is necessary that their SDK provides binding functionalities with VB.NET. Once a device has been enabled through its SDK, the Interfaces provided by the framework have to be used in order to implement the desired handlers and actions.

The built-in virtual environment of the framework, is able to load image files for the representation of products as well as flash files that can be used to represent monitor advertisements, digital price tags or other useful effects by just dragging and dropping them at any position in the racks. This environment was implemented in a way that makes modifications easy, so that it can also be used in other domains by loading different image backgrounds and audiovisual effects (flash files) without the need for additional implementations. This was possible by modifying and extending the PictureBox class provided by VB.NET. Its functionality was extended so that the elements (i.e., images and flash objects) loaded using an instance of this class can be freely moved on the area specified for the virtual environment, resized as well as register interaction events (e.g., gaze on object).

The interface responsible for the interaction patterns has access to the data structures of the virtual environment (i.e., flash and image objects). The methods provided by the interface can be used to design an interaction pattern. Based on the events that are triggered on the objects of the virtual environment, methods can be used to control the flash objects (e.g., start, pause and stop effect method) as well as

DRAFT

the image objects. For instance, it is possible to specify a sequence of method calls (e.g., start and stop calls) in this interface, so that if a gaze fixation occurs on a specific product image, a flash object will behave accordingly. In general, this interface is responsible for mapping an input to an output action.

In order to integrate an input device, the methods of the interface provided by the Input Media Library can be used to initialize and calibrate the input device as well as to start and stop tracking. The data streams (i.e., screen coordinate streams) of the input modalities are received by a method of the interface where also multimodal fusion can be applied. The methods of this interface currently support the integration of pointing devices. For other input types, such as speech input, the current methods have to be extended.

4 Evaluation

At the beginning of the paper we argued that knowing where users pay attention by tracking their gaze can enable novel interaction dialogues within future pervasive environments. Therefore our aim was to test with our viGaze framework how well the users attention can be guided or steered, while looking at an instrumented shelf in a pervasive retail environment. We performed a user study within the virtual retail environment to investigate whether the interaction examples designed using the *Interaction Library* can be realized. The following two interactions were implemented:

Interaction 1: *If a gaze fixation occurs on a product in the shelf for more than a period of 450 ms (this threshold was specified through a pilot study (see section 4.1)), the shelf triggers an advertisement for that product with a length of 8 sec. (a typical teaser video in a retail environment) on the display. If the monitor displays an advertisement, a fixation on another product does not change it.*

Interaction 2: *When the users attention shifts to the display for more than 450 ms, the users gaze is redirected to the product by letting the price tag under the product blink blue. If the user then again fixates on this product, the blinking effect stops.*

We define that the first part of the interaction cycle is completed, when interaction 1 is successfully triggered and the user fixates on the advertisement monitor. The second half of the cycle is completed, when the user's gaze fixates back to the product that started the monitor advertisement. If both interactions are completed after each other, then the interaction cycle is completed.

The aim of the defined interaction pattern was to investigate and explore the possibility of persuading users to follow the defined attention patterns in order to provide the product information on a monitor display that causes attention switching and finally redirect the attention back to the product (see figure 5).

On the one hand, when a user fixated on a product, we interpreted this behavior as interest and tried to provide the user with additional information about the product of interest by switching her attention to the monitor advertisement. On the other hand, the customer should not lose focus of the product. Thus, after watching the advertisement we used the flashing price tag effects to direct her attention back to the product.

4.1 Pilot Study

During a pilot study we evaluated the proposed effects (price tags flashing blue) and gaze fixation threshold. In total, 8 users participated in the pilot study, 4 male and 4 female. The criterion for the price tag effect was to find an effect that can easily be noticed by the participants, but still does not disturb them. We compared a red and blue color flashing effect for the price tags. The results showed that the blue color flashing effect fulfilled our criteria. All participants were able to notice the flashing effect independent of the color, but seven participants felt the blue color flashing effect less disturbing than

DRAFT

the red one. According to [19], the mean fixation duration during visual search is 275 milliseconds. We used this value for the first interaction tests and increased it to 300, 350, 400, 450 and 500 milliseconds in order to select a fixation threshold as a trigger for the interaction pattern. Our intention was to avoid interaction loops, e.g., starting the interaction pattern immediately after a product has been fixated, leading us to choose 450 milliseconds as the threshold for the gaze fixations. The participants also rated, using a questionnaire, the virtual shelves with an average of 2.75 points out of 4 according to how realistic they were and stated that the monitor advertisements were attention grabbing. We do not claim that the selected color and gaze fixation threshold are the optimal values to use. These values worked the best for the given configuration and only between the tested colors and thresholds. Finding the optimal values concerning the most attention grabbing color and the fixation threshold was not in the scope of this work.

4.2 User Study

For the actual user study 36 participants were recruited in exchange of a small honorarium. The mean age of the users was 25 years, with 17 male and 19 female participants, which came from different countries in Europe, Asia, and America. They had different cultural backgrounds and were students in various programs (Computer Science, Biology, Literature, etc.). All participants were exposed to shelves with the same product groups, but in randomly chosen configurations.

4.2.1 Apparatus and materials

For the experiment, a 24" wide screen computer monitor was used with a resolution of 1900 x 1080 pixels and the Tobii X60 eye tracker was utilized to capture gaze data at 60 Hz. The eye tracker and the speakers were placed under the computer screen that was located at about 60 cm from the participants.

4.2.2 Tasks and Procedure

For the study we created three different virtual shelf types with products from one of the following categories: cereals, soft drinks, or snacks (see figure 4). The participants were asked to perform the following tasks:

- **Task 1:** Explore the shelf.
- **Task 2:** Choose three products of your choice.
- **Task 3:** Find the three cheapest products.
- **Task 4:** Find the three products shown to you (The products were previously shown by the experimenter to the users on printed A4 paper).

The tasks were performed in order either from task 1 to task 4 or from task 4 to task 1. For instance, a user could be assigned to perform the tasks on the shelves of type 1 in the order 1 to 4, and on shelves of type 2 and 3 in the order 4 to 1. The purpose of these simple tasks was to investigate to what extent it is possible to influence the user's attention. The tasks were chosen so that the intrinsic attention control was increasing. For instance, in task 1 the participant is free in the decision where she wants to look at, whereas the aim of task 4 is very strict, because the participant must look for special products.

Participants were invited to take part in a study of a virtual shopping tour and were instructed that their eye movements would be recorded. There was no mention of the defined interactions and the purpose of the study. The experimenter read the first task to the participant and after that the first virtual shelf was shown to the participant where she had to perform the task and tell the outcome after 1 minute. Subsequently, the experimenter read the second task and showed the same shelf to the participant. This procedure was repeated until each task was performed on each of the three shelf types. Finally, after the study was over (~ 45 minutes), the participants had to complete a questionnaire in order to rate the

DRAFT

shelves and tasks as well as state the effects, interactions and behaviors that were noticed by them. In order to analyze the data, the successful interaction rates and a repeated measure analysis with factors shelf and task as well as a followed contrast between the single shelves were computed.

4.3 Results

The results of the study were retrieved from log files that stored the interaction sequences of each participant. We computed the results for each of the two interactions mentioned above, as well as the combined results. With occurrence success rate we will denote the successfully occurred interactions. For instance, if a user performs the first or the second half of the cycle then this is measured as a successful interaction. The analyses of these data revealed that the occurrence success rate of the first half of the cycle was 74.5%. The conditional probability (condition that the first half of the cycle actually occurred) of complete cycle occurrences was 42% and the total probability 31.1% (can also be seen as the second half of the cycle).

The success rate of the interaction cycle (interaction 1 and 2) was not significantly different between the tasks ($F = .446$, $p = .708$) (see Table 1). This means that independent of the task order (1-4 vs. 4-1) and so of "intrinsic motivation" the participant could be influenced in the same manner. In the case of the shelf factor there is a significant difference ($F = 5.576$, $p \leq .05$), (see Table 2). We evaluated the tasks on three different shelves. The results (amount of successful interaction switches) were significantly different only between shelf 1 and shelf 2. There was no significant difference between shelf 1 and 3 (see Table 2).

Table 1 Multivariate Test^b.

| Effect | | Value | F | Hypothesis df | Error df | Sig. |
|--------|----------------|-------|--------------------|---------------|----------|------|
| Shelf | Pillai's Trace | .247 | 5.576 ^a | 2.000 | 34.000 | .008 |
| Task | Pillai's Trace | .041 | .446 ^a | 3.000 | 33.000 | .708 |

a. Exact statistics, b. Design: Intercept

Table 2 Tests of Within-Subjects Contrasts.

| Source | Task | Type III Sum of Squares | df | Mean Square | F | Sig. |
|--------|---------------------|-------------------------|----|-------------|-------|------|
| Shelf | Shelf 2 vs. Shelf 1 | .259 | 1 | .259 | 8.419 | .006 |
| | Shelf 3 vs. Shelf 1 | .008 | 1 | .008 | .181 | .673 |

The participants were also given a short questionnaire after the study was over and were debriefed. They had to rate the shelves and tasks and provide some information about what they noticed. The questions were the following:

- **Q1.** How realistic were the shelves? (1 worst - 4 best)
- **Q2.** How realistic were the tasks? (1 worst - 4 best)
- **Q3.** Did you notice a systematic change of the advertisement display?
- **Q4.** Did you notice an effect under the product?
- **Q5.** Did you notice an interaction cycle (Product Monitor Product)?

The results retrieved from the questionnaires showed that the quality of the virtual shelves was rated with an average of 2.81 (SD = .18) and the tasks were rated with an average of 3.75 (SD = .10). Every participant except for one noticed the flashing effect under the products position, whereas the trigger for the effect was noticed only by 5 users (13.8%). The rules that caused the systematic change of the monitor advertisements were noticed by 17 users (47%) and finally the interaction cycles were noticed by only 5 users (13.8%).

DRAFT

5 Discussion

The results of this study demonstrated that it is possible to achieve the targeted user behavior by enabling appropriate interaction patterns and effects. On the one hand, the experiment revealed that it is possible to steer the user's attention by utilizing information gathered based on gaze fixations and on the other hand, it was possible through the experiment to point out the parts of the designed interaction pattern that can be optimized. The first half of the interaction pattern cycle was more successful than the second half. A possible reason for that could be the time interval in which the second half had to be completed (the time interval for the completion of the complete cycle was 8 seconds) or even that the effect that was enabled under the products position was only visual. A possible solution in order to increase the success rate of the second half of the cycle could be the addition of an audio stimulus. From this result we can infer that there is a lot of space for optimization that can be done in the second half of the cycle in order to increase the occurrence success rate of the complete cycle. The first half of the interaction consisted of a visual stimulus (video on monitor) combined with an audio stimulus (the shown video also had sound). Maybe this was also one reason (apart from time) why the second half was not as successful.

Although the tasks were chosen so that the intrinsic attention control was increasing, the results showed that the differences were not significant. This could be interpreted as a success for the defined interaction pattern since we managed to control and direct the users attention to the regions we wanted, independent of the task. On the other hand a possible reason could also be, that the attention grabbing effects that were chosen for the study were too obtrusive, forcing the user's attention to switch to the desired regions. Another possible reason could also be that the chosen time (1 min.) that was given to solve the tasks was too much, resulting in shelf exploration after each task was solved.

The differences between the first and third shelf were not significant, leading us to the conclusion that there was no learning process present. On the other hand, the success rate of cycle occurrences in the second tested shelf was less than in the other two shelves, opening the question whether a cyclic learning process was present.

One of the novelties of this framework is the integration of all necessary components into one system. In contrast to other approaches that have to use several software systems in order to be able to achieve similar functionalities, the proposed framework provides built-in methods for the design of a virtual environment for the performance of user studies. This has the advantage that implementation overload and the total time necessary from design to actual performance of a user study is reduced.

Through the performed user study we gained important insights about the functionalities provided by the framework as well as the effectiveness of testing. In only 1 hour, we were able to create 12 different shelves and enhance them with visual and audio effects. In total, we performed 432 trials (~27 hours) investigating interaction patterns to achieve the targeted user behavior. This number of trials is probably not possible to be reached in a real environment within such a time frame and costs, showing us once again the immense importance of such a virtual environment.

6 Conclusions

We introduced an eye tracking framework, an approach that allows a dynamic creation and design of virtual shelves, the enhancement of them with audio and visual information as well as the design and enablement of gaze-based interactions (explicit and implicit) that can take place between the users and the virtual shelves. Using a prototype implementation of the framework, we conducted a user study with 36 participants in order to investigate whether gaze can be used as an input modality to steer users attention in future pervasive environments. We also briefly highlighted that our framework is usable in different contexts, ranging from HCI to consumer behavior research. The presented framework viGaze can be considered a first step to integrate gaze-based interactions into instrumented environments. We have demonstrated its feasibility in the context of an instrumented retail environment, but believe that the ideas generalize easily to different kinds of instrumented environments. The full potential of viGaze will become obvious once eye tracking technology has matured to such an extent that it can be easily

DRAFT

installed into the environment, allowing for constant and calibration-free tracking of multiple users gazes. Until then, viGaze is a useful tool to explore novel gaze-based interaction concepts in a virtual mockup of an existing instrumented environment.

References

1. BIEDERT, R., BUSCHER, G., SCHWARZ, S., HEES, J., AND DENGEL, A. Text 2.0. In *CHI '10 extended abstracts on human factors in computing systems* (New York, NY, USA, 2010), ACM Press, pp. 4003–4008.
2. BULLING, A., ROGGEN, D., AND TRÖSTER, G. Eyemote—towards context-aware gaming using eye movements recorded from wearable electrooculography. *Fun and games* (2008), 33–45.
3. CAUCHARD, J. R., LÖCHTEFELD, M., IRANI, P., SCHOENING, J., KRÜGER, A., FRASER, M., AND SUBRAMANIAN, S. Visual separation in mobile multi-display environments. In *Proceedings of the 24th annual ACM symposium on User interface software and technology* (2011), ACM, pp. 451–460.
4. CORBETTA, M., AND SHULMAN, G. L. Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience* 3, 3 (2002), 201–215.
5. DAIBER, F., SCHÖNING, J., AND KRÜGER, A. Whole body interaction with geospatial data. In *Smart Graphics* (2009), Springer, pp. 81–92.
6. DUCHOWSKI, A. *Eye tracking methodology: Theory and practice*, vol. 373. Springer, 2007.
7. DUGGAN, G. B., AND PAYNE, S. J. Skim reading by satisficing: evidence from eye tracking. In *Proceedings of the 2011 annual conference on Human factors in computing systems* (2011), CHI '11, ACM, pp. 1141–1150.
8. FERSCHA, A., RESMERITA, S., AND HOLZMANN, C. Human computer confluence. *Universal Access in Ambient Intelligence Environments* (2007), 14–27.
9. GIANNOPOULOS, I., KIEFER, P., AND RAUBAL, M. Geogazemarks: providing gaze history for the orientation on small display maps. In *Proceedings of the 14th ACM international conference on Multimodal interaction* (New York, NY, USA, 2012), ICMI '12, ACM, pp. 165–172.
10. HEMPEL, J. Navigation with gaze and feet: An interaction method for spatial data. Faculty of Computer Science, University of Magdeburg, Bachelor Thesis, 2011.
11. HILL, R. L., DICKINSON, A., ARNOTT, J. L., GREGOR, P., AND MCIVER, L. Older web users' eye movements: experience counts. In *Proceedings of the 2011 annual conference on Human factors in computing systems* (2011), CHI '11, ACM, pp. 1151–1160.
12. JACOB, R. J. K. The use of eye movements in human-computer interaction techniques: What you look at is what you get. *ACM Trans. Inf. Syst.* 9, 2 (1991), 152–169.
13. KERN, D., MARSHALL, P., AND SCHMIDT, A. Gazemarks: gaze-based visual placeholders to ease attention switching. In *CHI* (2010), pp. 2093–2102.
14. KOMOGORTSEV, O., HOLLAND, C., AND CAMOU, J. Adaptive eye-gaze-guided interfaces: design & performance evaluation. In *CHI Extended Abstracts* (2011), pp. 1255–1260.
15. KRÜGER, A., SCHÖNING, J., AND OLIVIER, P. How computing will change the face of retail. *Computer* (2011), 84–87.
16. OHNO, T., MUKAWA, N., AND KAWATO, S. Just blink your eyes: A head-free gaze tracking system. In *CHI'03 extended abstracts on Human factors in computing systems* (2003), ACM, pp. 950–957.
17. PIETERS, R., AND WARLOP, L. Visual attention during brand choice: The impact of time pressure and task motivation. *International Journal of Research in Marketing* 16, 1 (1999), 1–16.
18. PORTA, M., RAVARELLI, A., AND SPAGNOLI, G. *eCursor*, a contextual eye cursor for general pointing in windows environments. In *ETRA* (2010), pp. 331–337.
19. RAYNER, K. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin* 124, 3 (Nov. 1998), 372–422.
20. RENSINK, R. A., O'REGAN, J. K., AND CLARK, J. J. To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science* 8, 5 (1997), 368–373.
21. SANTELLA, A., AGRAWALA, M., DECARLO, D., SALESIN, D., AND COHEN, M. Gaze-based interaction for semi-automatic photo cropping. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (2006), ACM, pp. 771–780.
22. SPASSOVA, L., SCHÖNING, J., KAHL, G., AND KRÜGER, A. Innovative retail laboratory. In *Roots for the Future of Ambient Intelligence. European Conference on Ambient Intelligence (AmI-09), 3rd, November* (2009), pp. 18–21.
23. STELLMACH, S., AND DACHSELT, R. Investigating gaze-supported multimodal pan and zoom. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (2012), ACM, pp. 357–360.
24. TANRIVERDI, V., AND JACOB, R. J. K. Interacting with eye movements in virtual environments. In *CHI* (2000), pp. 265–272.
25. TREUE, S. Visual attention: the where, what, how and why of saliency. *Curr Opin Neurobiol* 13, 4 (Aug. 2003), 428–432.

DRAFT



Fig. 1 The Innovative Retail Lab, an example for a future retail environment. See [15] for a detailed description of the various prototypes depicted.

DRAFT

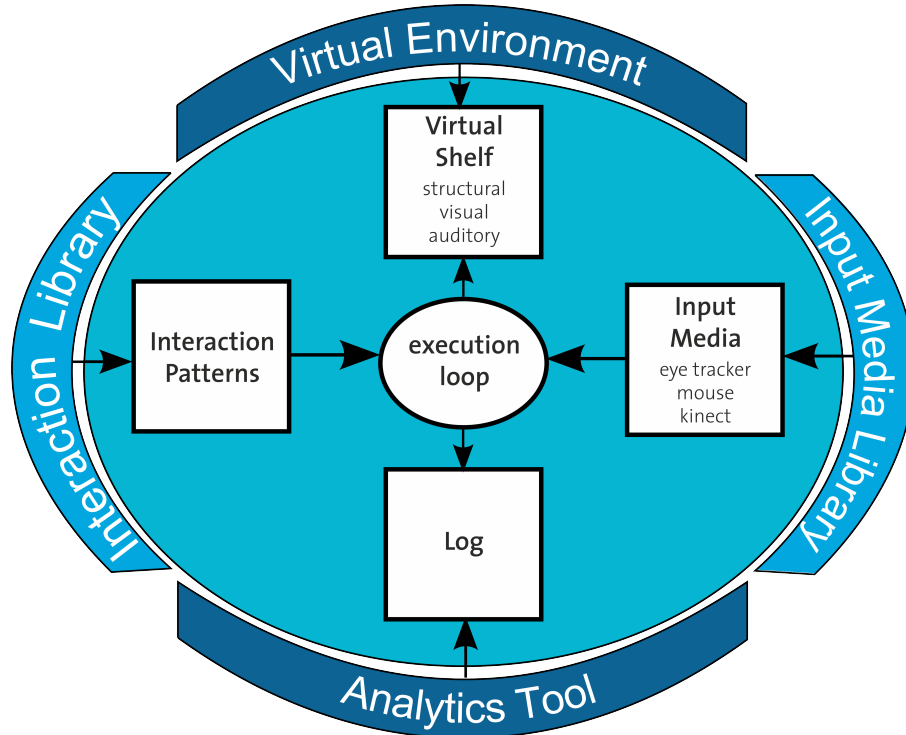


Fig. 2 The *Virtual Environment* component (top) is used to easily integrate various virtual pervasive environments into the framework. The *Input Medium Library* component (right) can be used to integrate a range of input devices such as an eye tracker or other Post-WIMP input devices. In the *Interaction Library* component (left), the interaction and relationship between the various sensors and actuators in the virtual pervasive environments can be defined. The *Analytics Tool* component (bottom) can be used to easily run and evaluate user studies to test the effectiveness of the interaction pattern and the paths defined in the *Interaction Library*.

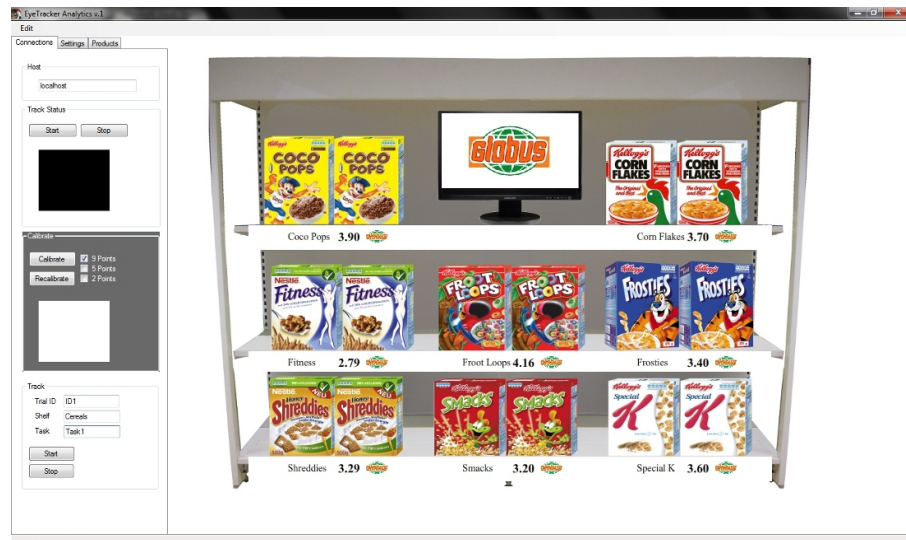


Fig. 3 User interface of the framework. The left side of the interface provides controls to design an environment, choose input methods and start a user study. The right side displays the actual virtual environment where the user can drag and drop the elements.

DRAFT



Fig. 4 Several configurations of a shelf with cereal products with different arrangements.



Fig. 5 When a gaze fixation on a product occurs, the advertisement video of this product is shown on the monitor. When a gaze fixation on the monitor advertisement occurs, a flashing effect starts at the position of the advertised product to redirect attention back to the product.

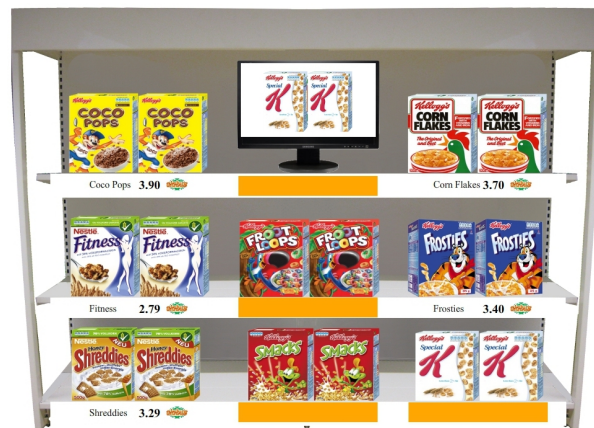


Fig. 6 When a gaze fixation on occurs on the advertisement monitor, the path to the product that is displayed can be highlighted.

DRAFT