LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN
Department "Institut für Informatik"
Lehr- und Forschungseinheit Medieninformatik
Prof. Dr. Heinrich Hußmann

Master Thesis

Pursuits-based Gaze Interaction using Real World Elements as Stimuli

Katharina Sachmann
Katharina.Sachmann@gmx.de

Bearbeitungszeitraum:	30. 01. 2017 bis 14.08.2017
Betreuer:	Mohamed Khamis
Verantw. Hochschullehrer:	Prof. Dr. Florian Alt

Zusammenfassung

Die Verwendung von Blick-basierten Konzepten erfreut sich immer größerer Beliebtheit für die Mensch-Maschine-Interaktion. Besonders häufig werden diese dabei für Überbrückung von Distanzen zum Ziel während der Kommunikation oder bei freihändiger Kommunikation mit öffentlichen Informationsanzeigen (public displays) eingesetzt. Sowie bei Geräten als Bestandteil von intelligenten Häusern (smart homes) oder Smart Watches. Mit der Einführung von Pursuits, wurde eine Technik, die eine Augen-Folgebewegung (smooth Pursuits) zu Nutze macht, vorgestellt. Gleichzeitig macht diese ein Kalibrieren vor dem Arbeiten mit Eye Trackern unnötig. Was die Technik sehr ansprechend für spontane und nutzerunabhängige Interaktion macht. Viele der bisher vorgestellten Anwendung auf diesem Bereich, nutzen Pursuits ausschließlich für Stimuli, die digital auf einem Bildschirm angezeigt werden. Nur wenige Forschung richtet sich auf die Anwendung von Pursuits mit Elementen der echten Welt. Hierbei wurden Objekte aber immer manipuliert, so dass ein Stimulus vorhanden war, welcher eine Augenbewegung initialisiert.

Im Rahmen dieser Arbeit wird das Potential und die Machbarkeit von Pursuits-basierter Blick Interaktion mit Elementen aus der echten Welt als Stimulus untersucht. Dafür vergleichen wir die von Pursuits erzeugten Ergebnisse beim Betrachten von Zielen in der echten Welt, vorgeführt in verschiedenen Umgebungen: (a) die Situation in der echten Welt, (b) eine unbearbeitete Videoaufzeichnung am Monitor und (c) eine bearbeitete Version des Videos, aus dem alle Tiefeneindrücke entfernt wurden. Basierend auf den Ergebnissen dieses Vergleichs wurde eine Nutzerstudie in der realen Welt entworfen, mit der die Akzeptanz der Nutzer und die technische Leistung evaluiert wurden. Die Ergebnisse lassen darauf schließen, dass das Verwenden von natürlichen Folgebewegungen als eine positive Interaktionsart von den Nutzern empfunden wird.

Abstract

Using gaze is a popular way to design concepts for Human-Computer-Interaction (HCI). It is especially often used for remote or hands-free communication with for example public displays, devices within smart homes or smart watches. With the introduction of Pursuits, a new technique, that exploits the smooth pursuit eye movement, was presented. It overcomes the need for calibration before working with an eye tracker, making it very attractive for spontaneous and user-independent interaction. The majority of applications currently focus on the adaption of Pursuits for moving on-screen targets and only very little research is directed at the use of Pursuits for real world target. However here usually the object is altered and animated to display a smooth pursuit stimulating movement.

In this work we explore the potential and feasibility of Pursuits-based gaze interaction with moving stimuli in the form of real world elements. Therefore we compare the performance of Pursuits resulting from pursuing a real world target in different environments: (a) the real world situation, (b) the same situation shown in the form of an unedited video and (c) an abstracted video version dismissing potential depths cues, the two latter respectively on-screen. Based on the results of this experiment we designed a usability study in the real world, where we tested three scenarios on user acceptance and Pursuits performance. Gathering that exploiting natural smooth pursuits movements is perceived as a positive interaction method by the user.

Task Description

Leveraging the smooth pursuit eye movements for gaze-based interaction (aka Pursuits) is increasingly becoming popular. Pursuits has been employed for calibration-free gaze interaction with public displays, smart watches, and smart homes. A key advantage of pursuits, is that it is a natural movement that humans perform unconsciously when following a moving object. At the same time, we are living in dynamic environments and are continuously surrounded by moving objects, ranging from vehicles, other humans, and also objects that move from our perspective as a result of moving closer or farther away from them (e.g., we perceive trees to be approaching us while we are driving on a highway).

The goal of this thesis is to explore the potential and feasibility of gaze-based interaction by utilizing the already-moving objects that surround us in our daily lives. A sample scenario is a situation where a user is in a restaurant and wants to call the waiter; the user can pursue the waiter with her eyes, the system would then detect the smooth pursuit eye movement and notify the waiter that he is needed.

In a first step, a comparison has to be made to investigate how Pursuits selections in the real world differ from those traditionally done on computer monitors. This can be done through a study in which selection of real objects in the real world is compared with selection of the same objects in video recordings. Depending on the results, the following studies would be conducted either in the lab by using video recordings of objects moving in the real world, or by recreating a realistic scenario where participants select an object directly in the real world.

Tasks:

- Review of related literature on gaze-based interaction using Pursuits.
- Conducting a study to investigate how Pursuits selections in the real world differ compared to when on-screen selections.
- Investigating use cases for Pursuits in the real world.
- Evaluating some of the use cases to understand how the concept is perceived by users.

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbstständig angefertigt, alle Zitate als solche kenntlich gemacht sowie alle benutzten Quellen und Hilfsmittel angegeben habe.

München, 14. August 2017

. . . .

Contents

1 Introduction

Would it not be great to just look at something around you and be able to interact with it. No mouse, no pointer, no touch, no mid-air gestures, no speaking. Just looking at it. Imagine sitting in a restaurant and trying to get the bill for the last twenty minutes but the waiter seems too busy to even notice you. But now just by looking at them you could let them know that your table is ready to leave - at no extra cost. With Pursuits-based gaze interaction using real world stimuli we suggest applying the motion-based interaction method Pursuits to real world target. Enabling users to interact with moving objects around them using their eyes.

Gaze is a powerful means in interaction and used in Human-Computer-Interaction (HCI) for a long time, first designs of gaze-interaction methods being introduced as early as 1981 [3, 20]. Besides selecting and moving objects, adapting information depending on the user's attentions or navigating the scene, it can also be used for gaming [16, 39, 40] or more complex tasks like typing text on a specially designed gaze keyboard [12]. The developed applications often not only rely on dwell time or utilizing gaze gestures [9]. But also combine it with other input modalities such as touch on the screen directly [32] or a hand-held remote [38]. Making gaze interaction in general more stable and extend interaction scenarios, the same way Pursuits does for the spontaneous interaction with public displays [45].

Pursuits is an interaction method based on the eye movement smooth pursuits, which occurs if the human eye constantly follows a moving stimulus. Its advantage is, that the smooth movement can hardly be faked without a visual stimulus. Providing a good property to identify matching eye and object movements with almost no doubt. By collecting data of the user's gaze as well as the the stimulus over for a while, a movement log for both can be created. Periodically the correlation between both of these logs is calculated stating a possibility that the eye movement is due to the identified object movement. By correlating two movements relatively to each other the sometimes tedious process of calibration can be omitted. Pursuits is well-established for the use on (public) displays either to enable direct interaction [10, 26] or ease calibration [33].

An attempt was made to enable non-screen Pursuit-based interaction with AmbiGaze [42] applying Pursuits to a smart environment where devices were alter in such a way that they display movements that can be smoothly pursued. With Pursuit-base gaze interaction with real world elements as stimuli we suggest using the omnipresent movements of objects around us as stimuli for the smooth pursuit movement in order for Pursuits to work.

Following this introduction, section 2 Background will provide a short Background beginning with an overview of commonly used eye movements in HCI and their application in section 2.1 Eye movement Types in Human-Computer-Interaction. It will then provide an introduction to eye tracking techniques in section 2.2 Tracking Techniques and some applications of gaze interaction in public in section 2.3 Gaze Interaction in Public. We reasoning our intention of applying Pursuits to real world objects in section 2.4 Summary and Motivation.

In the first main part of this work, section 3 Eye movement Types in Human-Computer-Interaction, we explore Pursuits in the real world. By firstly investigating difference between aspects of eye movement reacting to a 2D on-screen and a 3D real world displayed in section 3.1 Following Movement in the Real World vs. on-screen - Motivation. After that we describe the set up of our pre-study in section 3.2 Pre-Study: Analysing Pursuits in different Environments, the evaluation process in section 3.3 Pre-Study: Evaluation and the study results in section 3.4 Pre-Study: Results. We close this section by highlighting our lessons learnt in section 3.5 Pre-Study: Results.

The second main part focuses on the development and testing of use case in section 4 Developing and Testing Use Cases. In section 4.1 Brainstorming Session we firstly display the design space resulting from a focus group, which was held to come up with use cases. The research questions implied by that design space are discussed in section 4.2 Design Space - Research Questions. Before we explain the set up of our usability study in 4.3 Usability Study and show its evaluation and results in section 4.4 Evaluation & Results. We close with a discussion of the results in section 5 Discussion, ideas for future work in section 6 Future Work and a conclusion in section 7 Conclusion.

2 Background

This section presents an overview of necessary background to the subject of eye tracking and gaze-based interaction. Introducing eye movements that are often used in human-computer-Interaction joined by the respective interaction method, before illustrating different eye tracking techniques. It closes with a couple of works on gaze interaction in public and a summarising motivation.

2.1 Eye movement Types in Human-Computer-Interaction

The eye's muscles performs several different movements (sometimes at once). Some as small as the adaptation of the lenses curvature other as noticeable as the rolling of the eyes. Only few of them are applicable for human-computer interaction (see e.g. [21]). In the following we present the most common eye movements used for interacting with computer-systems: *fixations*, *saccades* and *smooth pursuit*. Each part will include an introduction to their matching interaction method.

2.1.1 Fixations - Dwell Time

A fixation is the pause the eye takes in order to perceive information within the small area of high acuity. This pause usually takes between 200ms to 600ms and covers an area of one degree visual angle. During this time the eye is mostly still except for omni-present jittering of less than one degree of visual angle [20]. Depending on the accuracy of the used eye tracker, the distance to the device and the target size on the screen, this jittery can aggravate tracking. In order to use fixations as a means for interaction and overcome the Midas touch *dwell time* is used. Midas touch is the case of accidentally selecting everything that is looked at due to the recognition of a gaze point at this position. By purposely dwelling on the desired target for a longer period of time, that last longer than a usual fixation would, a deliberation is communicated.

This starring is presumed and reported as unnatural [11, 21] tough. Besides also being assumed to not being able to keep up with other selection methods such as the press of a button. As mechanism like this can be trained and performance improved over time, resulting in shorter completion times. Whereas the use of dwell time 'provides a permanent barrier against [...] speed-up'[48]. This method and the delay is especially impractical where fast sequences of or parallel actions are necessary such as when used during gaming [17, 19]. Here, using fixations and dwell time, is unsuitable as selections have to made quickly or in parallel. Also because targets are often in motion or small which make fixations impossible. Additionally some games require the selection of labelled targets. In order to read a word, the eye has to fixate on it for a longer time, that usual leads to the Midas touch effect of wrong selections [19]. Which can also occur in other context such as typing with a adapted keyboard, that offers suggestions. Before a conscious choice can be made the text needs to be read by the user which takes longer than the dwell time and thereby triggers an action [13]. But a combination of fixation and manual interaction, dismissing the dwell time, turns out to not be faster than a dwell time-based approach when entering short strings like a numerical PIN [8].

2.1.2 Saccades - Gaze Gestures

To perceive and see a scene in its entirety the eyes needs to scan over the the area. This is archived with so-called *saccades*. These are very quick jumps of the eye, usually done in 30ms to 120ms, and bypassing an area lager than two degrees of visual angle [20, 21, 29]. These saccades can be used in the form of gaze gestures, which are 'a controlled saccade or series of controlled saccades that causes an action when completed' [31]. Using gestures instead of dwell time intends to not only reduce delays resulting from dwell time but also overcome the Midas touch problem, accuracy issues and calibration shifts. As the latter is not longer necessary because

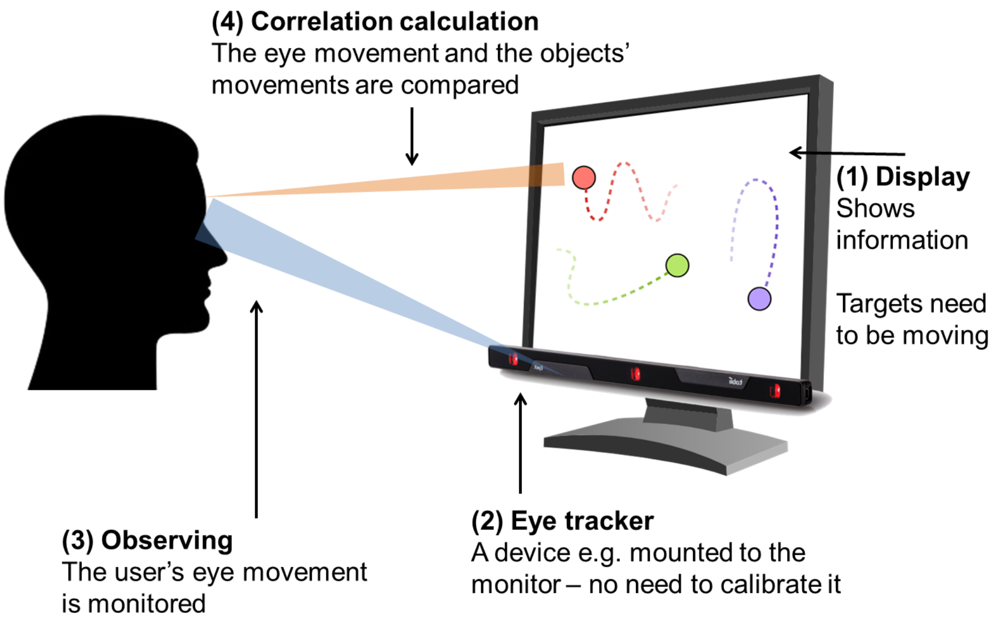

(4) Correlation calculation
The eye movement and the objects' movements are compared
(1) Display
Shows information
Targets need to be moving
(3) Observing
The user's eye movement is monitored
(2) Eye tracker
A device e.g. mounted to the monitor – no need to calibrate it

Figure 2.1: Components and process of Pursuits. [Altered diagram of [45]]

trajectories between start and end fixation are considered and no longer exact points on screen. This is also why accuracy of the interaction is improved and the Midas touch no longer an issue [9].

A saccade is a *stroke* that starts and ends with a fixation. A gesture can consist of several saccades, including several fixations in order to determine the shape of the gesture [15]. This leads to an expected minimal execution time that sums up the duration of all saccades and needed fixations. When tested however this time is rarely archived by the users [15, 31] as they would often use more fixations, interrupting a saccade, than necessary in order to execute a gesture. But tests showed that this time is at least not depended on the scale of execution. Meaning the speed increases and one saccade covers a bigger distance compared to the same gesture in a smaller scale [15]. Nor is it longer than the duration of an action applying dwell-time interaction [18]. Is has also been demonstrated that the time needed to perform a gaze gesture is within the same range of the execution time that is needed for the same action being completed with mouse and keyboard [9].

2.1.3 Smooth Pursuit - Pursuits

The a rather new interaction method is utilizing the smooth pursuit movement. This movement is triggered by a moving stimulus, in order to keep that in focus the eye pursues the object. Smooth pursuits are hard to fake as they cannot be performed without a stimulus producing a movement [29]. Is was used for example to gain information about a user's conditions [44] before it was introduced as an interaction technique in 2013 [45]. It overcomes the need for calibration as well as accuracy issues and the Midas touch problem, for the same reason stated above (no need for exact gaze location). The changing position of the target as well as the gaze position have to monitored over time and are periodically compared to each other, calculating the Pearson's product-moment of correlation to determine the degree of match between both (see fig. 2.1). The longer this recording period (time window) is, the more information it contains about the overall direction of either gaze or object, especially for slow movements. But this also means that the delay between the start of interaction and response is further delayed.

As Pursuits does not rely on a exact position but on matching trajectories it works very well also for small targets, enabling for example hands-free interaction with the small screen of smart watches [10]. Tests revealed that circular trajectories work best with larger diameters and that the stimulus' speed needs to be chosen carefully. Because fast targets lead to saccadic movements of the eye whereas slow speeds initiate fixations. Exploiting the fact that user's gaze dwells on moving targets is also a calibration method for eye trackers applying Pursuits [33]. This way calibration cannot only be accomplished more playful but also without the often as unpleasant perceived fixations [21, 33].

A use case that was already presented for the application of dwell time can also be approached with Pursuits: Entering text by Pursuits [28]. Pursuits is also often applied for interacting with larger displays or objects further away. An introduction to some of these approaches is presented separately in section 2.3 Gaze Interaction in Public and a more detailed overview can be found in [41].

2.1.4 Optokinetic Nystagmus and Vestibulo Ocular Reflex

An eye movement that is know to us all even if we cannot name it is *optokinetic nystagmus* [21]. That is the very shaky movement the eye performs to stabilise the object of interest the perceived images (e.g. when looking at a fast passing train). In order to do so the eye performs a combination of saccades and smooth pursuits. By detecting the difference in the length of the smooth pursuit phases and categorising them with machine learning techniques objects can be filtered that caught the users eye [22].

The vestibulo ocular reflex describes the adaptive movement of the eye balls when we move and turn our head. As these movements performed by the eye do not actually indicate change in gaze direction but in head direction, it is considered a non-visual movement. Opposed to all types described above, which always lead to a change of the perceived image. This way, using eye tracking, enables the use of head gestures (eye-based head gestures) without motion sensing devices.

2.2 Tracking Techniques

The information used by all approaches in section 2.1 Eye movement Types in Human-Computer-Interaction is obtained by so called eye trackers. These devices can either be mobile and be carried around (also referred to as head-mounted) or be stationary mounted to a display (also referred to as remote tracking). For both versions different techniques can be used such as the three predominate types: *video based* or *infra-red (video) based* tracking as well as an technique called *electrooculography*. These types are described in the following.

2.2.1 Video based tracking

In a video based approach off-the-shelf camera devices can be used to gather data [29]. A camera observing the eyes can either be used as a display mounted or a mobile tracking device attached to the user's head. The latter often also includes a front-facing camera to map the gaze point into the world. In order to process the data a computer is needed in both cases. Image processing is used to detect the position of the pupil. Without any reference points this data is only meaningful if the user keeps the head still. To not restrain users in their movement, reference points like a light reflection or the corners of the eyes are used.

Because video images are pixel based the distance between camera and eye highly influences the quality of tracking [29]. The further the camera is away the smaller the area which represents the eyes/pupils. This is why the resolution of the image plays an important role. In a remote scenario therefore cameras with a larger focal length are used, which comes with the tradeoff of a smaller camera view. Additionally choosing a camera with a high framerate improves accuracy since position changes can be detected quickly.

Besides the already mention requirements for the components, using a video based approach relies on a unobstructed view of the eyes, meaning that e.g. glasses can decrease the detection rate of the pupil. Furthermore constantly good light conditions are required and it does not work well on small eyes [29].

2.2.2 Infra-red based tracking

Without the comprehension with referent points in solely video image based tracking head movements render an accurate tracking almost impossible. Using infra-red (video) based tracking (IR-based) still applies the procession of video images but automatically adds an artificial infra-red light source. This light is invisible (and harmless) to humans but creates a corneal reflection that can be detected with special infra-red cameras. Depending on the angle of incidence the pupil will appear very bright (referred to as ‘bright pupil tracking’) or very dark in the image (referred to as ‘dark pupil tracking’) [29]. The further processing is identical to the video bases approach with reference points: the position of the pupil is observed in relation to the reflection which indicates the gaze direction of the user.

Quality of tracking can also be improved by using a camera with a high framerate and resolution. As IR-based tracking automatically provides a light source it is less sensitive too weak light conditions in-doors but the changing ambient light conditions when working e.g. outside influences performance of IR-tracking [2].

This technique can be applied to stationary trackers such as the Tobii EyeX[1] or mobile devices such as the Pupil tracker[2], which are used in the studies described in section 3 Exploring Pursuits in the Real World and 4 Developing and Testing Use Cases.

2.2.3 Electrooculography

The last introduced technique *electrooculography (EOG)* measures electrical signals at the eyes. As the human eye manifest a positive pole at the cornea, the front of the eye, and a negative pole at the retina, the back of the eye, a electrical field is created [29]. Eye movements lead to a new adjustment of this electrical field as the eye balls including the two poles rotate. This change can be detected with electrodes on both sides of the eye (e.g. temples and forehead). If one of the eye balls rotates towards the nose the positive pole is pointing towards the electrode placed on the forehead and the electrode placed on the respective temple detects the negative charge of the retina.

Both image based techniques are in their own way sensitive to light as EOG does not us light based data it is insensitive to light chances. However it comes with its own drawbacks. Nowadays we are surrounded by devices creating electrosmog, that are artificially created electrical fields. As EOG is based on electric signals it is prone to electrical noise in the same way video based approaches are to light [29]. Furthermore it ‘provides lower spacial POG [point of gaze] tracking accuracy’ than video and IR-based approaches and can used better for detecting gaze gestures.

[1] tobiigaming.com

[2] pupil-labs.com

In order for this to work the electrodes need to be attached to the user's face. Hence only a mobile, head-mounted version of this technique can be implemented.

2.3 Gaze Interaction in Public

Users often hesitate to interact with e.g. displays in public because they are afraid of embarrassment due to mistakes [4]. Gaze interaction offers a great means to interact over distances and without physical contact or visible action [45]. Therefore we focus on a couple of gaze applications for public interaction in the following section distinguishing between concepts for display and objects.

2.3.1 Interaction with Public Displays

Interaction not always requires an explicit input action by the user but can also be in the form of responding to behaviour. Gaze aware displays alter the displayed content automatically without an active user interaction [34]. Displays like these register attention of one or more user and enable them to interact parallel. A 'you are here' map for example reveals information about the nearby sights and facilities like hotels in the area, if the user looks at the position in the map. A video based tracking approach is sufficient to enable users to interact with gaze gestures or similar concepts [49]. With reserved regions on the display for indicating actions [31] the user can issue a command to the system. This approach seem especially interesting for content that can be flicked trough or binary choices. Pursuits is also a very promising interaction method for public display as calibration is not required. So demonstrated by several applications e.g. an information display, music players/libraries [45, 46], public authentication [7, 45] or entertainment [24, 25, 45]. As these systems all display graphical items as stimuli they exclude a big area of public displays, which show textual information. Systems displaying text have been tested, also in the wild, and results were generally promising [27, 26].

These outcomes suggest that applying Pursuits as an interaction mechanism for public displays is a welcome approach. But however not all information on public displays can be designed in such a way that objects or text float on screen. Therefore accurate and directed interaction is still necessary. This is why calibration is still important. But Pursuits overcomes the tedious and uninteresting fixation based calibration. As for calibration it is only necessary to know the position of the point currently looked at on the output screen. With a mapping of the trajectories like during Pursuits attention towards a target can be confirmed and gaze positions be mapped to certain points on screen passed during the object's movement [33]. Again research in the beginning covered mostly graphical stimuli or text labels that need to be focus on. However a vast number of screens display textual information, therefore attempts towards calibration based on the movements during reading were made [26]. Only revealing the text partially when reading the system guides the user's gaze in the same way a floating object would. Studies showed that calibrating displays this way is realistic even though less accurate than graphic or fixation based approaches.

2.3.2 Interaction with Things

Nowadays we are used to being able to interact with essentially everything, so why not use gaze to interact with everything around us [1]. An all-aware environment could be able to detect directed gaze and offer interaction. With eye contact sensors (ECS) objects can be equipped to detect when users are looking at them and respond accordingly. In a smart environment for example entertainment appliances can themselves determine when they are looked at and issue their sets of controls to a central remote control [43]. In order for this to work however a constant monitoring of the space and the user is required as the attention detection is controlled by the various devices and thereby privacy can be violated.

The constant observation of the user can be eliminated by transferring the active part of the system, that detects attention, to one single processing instance and the user. By animating the objects in the smart environments attention detection via the matching of gaze movement to the different motions displayed on the devices (Pursuits) is implemented [42]. The gadgets themselves do not need to constantly scan the area and privacy is maintained. However, in both cases, the objects in the environment have to altered and connected to an overall system in order to communicate their stimulus' position. Otherwise gaze and object position cannot be correlated and no interaction is possible.

2.4 Summary and Motivation

Gaze is a promising means of interaction providing several movements that can be exploited to issue commands or adapt context aware system. Three often used methods are firstly dwell time based on deliberate gaze *fixations*. Secondly gaze gesturing based on the quick and jumping eye movements *saccades* and lastly Pursuits utilizing the smooth pursuit eye movement. That is initiated when following an stimulus and can hardly be faked without said stimulus. These eye movements and produced gaze positions can be detected and recorded with different more or less invasive eye tracking techniques. The tracking can either be solely video based or infra-red (video) based, where videos of the eyes are recorded and image processing is used to find the position of the pupil. Based on that a gaze position on the target is calculate afterwards. Both techniques can be applied to either mobile (head-mounted) trackers that allow the user to walk around and potentially interact with everything in their surrounding. Or to stationary (remote) trackers that need to be connected to the target device and are only able to map gaze there.

The interaction mechanisms addressed above can be applied to several use cases and have strengths and weaknesses. One mutual strength is the possibility to interact over great distances and without physical contact to the target of interaction, making gaze an ideal modality to utilise for set-ups in public. Pursuits is a particularly interesting technique as it does not require a tedious calibration process beforehand, that is usually necessary in order to map measured gaze points onto the output space. The majority of concepts targets interaction with public displays with graphical contents and some with textual contents. Only few investigate the possibility of interacting with objects in the environment rather than especially designed context on a display. However the introduced approaches require an alteration of the environment in order to be able to interact with it. Either by making the elements of the environment aware of users or introduce an independent component, that can register and distinguish attentions toward single objects. The latter approach applies Pursuits to animated objects in the wild however interaction is only possible for devices connected to the overall system and not arbitrary objects entering or leaving the scene.

We suggest to close this gap by applying Pursuits to all objects in the real world, as our environments offers many moving stimuli such as cars, dishes in a sushi restaurant, animals or other people. Even stationary objects like buildings or street signs can move from a user perspective if users themselves are in motion. These movements can be exploited using Pursuits. Due to the lack of technical implementations needed to collect all necessary data about possible targets within the environment (e.g. image processing on video data of the front-facing camera) we begin by investigating performance and user acceptance of the interaction method itself.

3 Exploring Pursuits in the Real World

Before we can design and test concepts applicable to use cases for Pursuits-based interaction we need to investigate gaze behaviour in three dimensional space. More specifically the performance of Pursuits following movements in three dimensional space of real world scenarios. A divergent rate of positive selections from results achieved with two dimensional on-screen movements can determine how to test real world Pursuits scenarios as well as reveal design aspects for such scenarios. In order to detect these differences we conducted a study comparing the performance of Pursuits with a stimulus moving in three dimensional real world and the same movements displayed on-screen. The motivation, the study design as well as the study itself along side the evaluation process and its results are discussed in the section below.

3.1 Following Movement in Real World vs. On-Screen - Motivation

The human has a binocular field of view of approximately 200° horizontally and 130° vertically, although the eye is only able to focus on a very small area of that field, humans are receptive to visual stimuli in that area [36]. If we use a video prototype to test real world Pursuits scenarios on-screen we would only provide the stimulus in a small area within the field of view. Since stimulus movements in three dimensional space can take place within the entire field, the eye movements made to follow a moving object on-screen might not be identical. Also trajectories of the stimulus might appear differently when put to the screen. Potentially also leading to different eye movements and thereby it might influence the performance of Pursuits. Imagine an object frontally approaching and then passing the on the right. On-screen that object is moving to the right and leaving as soon as the screen ends (e.g. once out of the cameras field of view). While in the real world the eye will not only react with sidewards movement but also other forms of adapting (see below) to the depth change. Because of the bigger field of view the user can follow that movement longer.

To be able to perceive movement as three dimensional the brain needs to able to detect depth. When looking at a (computer) screen we either are presented with a flat 2D display of information (applications like Microsoft Word, Browser etc.) or an artificially generated 3D-Scenes (games, models etc.) creating an impression of a space leading into the screen (increased by the use of e.g. stereoscopic film). To achieve this impression the eye is tricked by simulating certain depth cues. However not all of the natural depth cues can be manipulated when looking at a screen.

Visual depth cues can be provided when displaying two and three dimensional imagery for example on-screen. These include but are not limited to *relative size of the object*, *shadow and light distribution*, *occlusion* or *motion parallax* [35]. Whereas so called oculomotor depth cues (accommodations and convergence) can only be partially simulated (e.g. by using stereoscopic film). Accommodation describes the mechanism controlling the current focus point of the eye's lens by altering its curvature. Interplaying with convergence, which is the mechanism controlling the rotation of the eye balls. This provides slightly different images for each eye. When looking at a screen the eye's focus will always stay on the display even though the convergence of the eye balls might be differ (relevant for e.g. stereoscopic film) [35]. Visual depth and oculomotor cues can transport different information about the depth-factor of a motion and thereby might also influence the performance of Pursuits.

As one approach to test 'real world' Pursuits scenarios would be to re-create them on screen a study was designed to investigate whether the size of the area, the trajectories of the stimulus or the existence of different depth cues has an effect on the performance of Pursuits.

Overview: Environment Conditions		
Session one	Session two	
Real World	Original Videos	Abstract Video
• Mobile eye tracker • Stimulus moving in laboratory space • Recording gaze data & world video • Visual and oculomotor depth cues available	• Stationary eye tracker • Un-edited video of first session on-screen • Recording gaze data • Visual depth cues available	• Stationary eye tracker • Edited black and video of first session on-screen • Stimulus movement mapped to white circle, removing all depth cues • Recording gaze data

Table 3.1: Pre-Study:Three Environment Conditions (Independent variable 1)

3.2 Pre-Study: Analysing Pursuits in different Environments

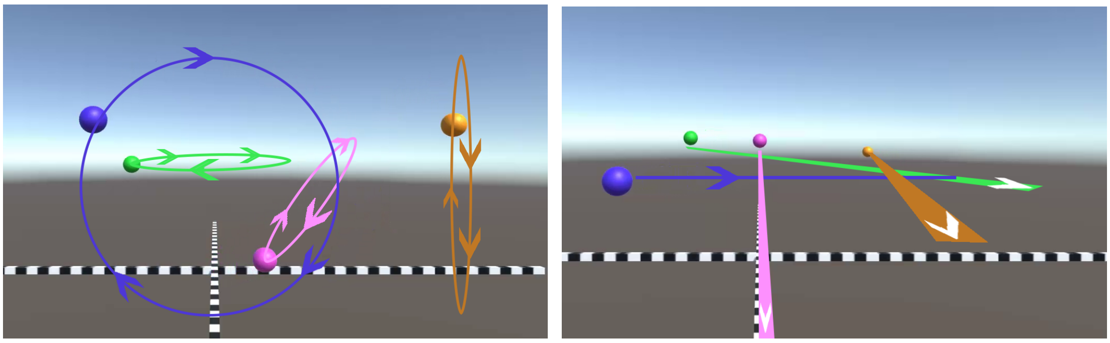
With a pre-study we want to investigate differences in performance of Pursuits in three different environment conditions on eight different movements types.

3.2.1 Pre-Study: Study Design

The study was designed as a within subject repeated measure experiment, where each participant covered all conditions. Over two sessions we changed three environmental conditions (see Table 3.1). In the first session the user was asked to follow a provided stimulus (red ball on a stick) moving in the laboratory space (see fig. 3.2(a)). During that process they wore a mobile eye tracker. Besides the gaze data we also recorded the video of the tracker's front facing camera, not only for later evaluation purposes but also as material for the additional session. The video was edited to only show the stimulus' movements mapped to a white circle on black background, neglecting all visual depth cues from the video (see fig. 3.2(c)). During the second session, which took place two weeks after the first, we showed the user the video recorded at their respective first session as once the original video without any changes and the once the abstract video in black and white with all depth cues removed (see fig. 3.2(b,c)). The abstract video was manually created with Adobe After Effects[3] by mapping the white circle to the stimulus if a movement condition was performed. We also recorded the gaze data during the second session.

During the first session the stimulus had to be moved by a researcher (see fig. **??**) along eight pre-defined trajectories, four circular and four linear trajectories (see Table 3.2). The movements are described within a right-handed coordinate system (x describing width, y height, z the depth). Both baseline movements are presented on a vertical plane in front of the user (xy-plane): In the circular case a circle was drawn in the air with the stimulus having the z-Axis acting as the circles centre. In the linear case a straight line was drawn parallel to the x-Axis. These movements do not have an depth aspect thereby being similar to the movements created on-screen for usage of Pursuits until now and providing us with a sort of baseline (see blue line in fig. 3.1).

[3] www.adobe.com/de/products/aftereffects.html



(a) Circular Movements (b) Linear Movements

Figure 3.1: Pre-Study: Set of trajectories used in the study visualised in a 3D-model

	Circular	Linear
Baseline	Circular Baseline	Linear Baseline
Alterations	C-Tilted	L-Height
	C-xAxis	L-DiagonalPass
	C-yAxis	L-zAxis

Table 3.2: Pre-Study: Eight Trajectories (Independent variable 2)

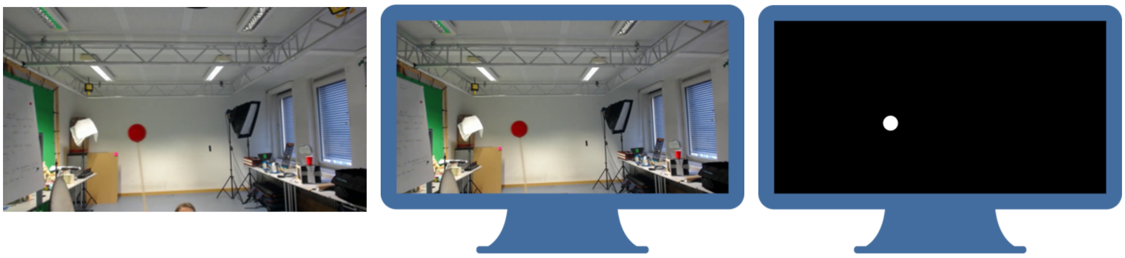
The remaining circular trajectories are circles rotated in the room so that x- and y-axis (C-xAxis, C-yAxis) both act as the centre of the circle (see brown,green circle in 3.1(a)). A fourth rotation was introduced tilting the circle by 45° in regards to all axis (C-Tilted, see pink circle in Fig. 3.1(a)). The remaining linear trajectories cover an object approaching frontally with changing height: movement along the z-axis while only changing the y-value (L-Height, see pink line in fig 3.1(b)). An object approaching at one side of the user: movement started in the negative x-segment and staying parallel to the z-axis, x/y-value remaining the same (L-zAxis, brown line in fig. 3.1(b)).And lastly an object approaching, starting in the negative x-segment and switch to the positive x-segment while approaching - or vice versa (movements along z-axis, switching x-segments), so that the object would pass the user in front (L-DiagonalPass, see green line in fig. 3.1(b)). The eight movement types were randomised using a balanced latin square (can be found in the appendix A).

All movements were performed three times (e.g. three circles or walking three times back and forth) in front of the user and on eye level (except for L-Height).

3.2.2 PreStudy: Procedure and Study Set-up

The study took place in university facilities. The first session was held in a room big enough to execute all trajectories whilst having the user stand at one end of the room looking along the z-Axis. After the eye tracker was calibrated with a 9-point on-screen calibration, the baseline conditions were performed. They were done approximately two metres in front of the user, as well as all the circular movements were centred there. Logically the distance for movements with depth aspect, especially the linear conditions, varied. During this session a mobile Pupil[4] eye tracker in their monocular form equipped with one eye camera and one front facing/world camera

[4] pupil-labs.com



(a) Real world (b) Orginal video (c) Abstract video

Figure 3.2: Pre-Study: The environment conditions as used in the pre-study

was used. Gaze is captured with a gaze accuracy of 0.6 degrees and a sampling rate of system limited 30Hz. The eye camera was only used to collect gaze data, no eye video was recorded. The front facing camera recorded the scene. This resulting video was shown to the user in their second session once without changes and once the abstracted version (both times they were shown in the same resolution as recorded). A laptop was placed on a table in front of the user in an smaller office room. Having them look at the display as frontally as possible (depending on the users height). A video-player implemented with C# on the .NET Framework 4.5.2 was used to show the participants the videos. It would not only display the video but log the gaze data from the moment play was hit until the video stopped. For implementing gaze logging the Tobii API was used as a stationary Tobii EyeX[5] eye tracker mounted to the laptop was used to collect gaze data with a sampling rate $\geq$ 60 Hz. The tracker was calibrated with at least six points. As a visual stimulus a home-made target in form of a red-painted polystyrene ball on a wooden stick (⌀20cm, length 50cm) was used, this was visible in the *real world* and *original video* environment.

3.3 Pre-Study: Evaluation

The evaluation of the data took place after the study and was developed over time. Including the establishing of evaluation aspects, a manual analysis in the beginning and the implementation of a evaluation tool.

3.3.1 Evaluation Aspects

In order to compare the performance of Pursuits we carried out analyses regarding the single environments before comparing them to each other. Therefore we investigated several aspects of the Pursuits algorithm:

(A) True positive rate The number of times the correlation between the participant's gaze data and object data exceeds a chosen threshold, calculated per participant in percentage

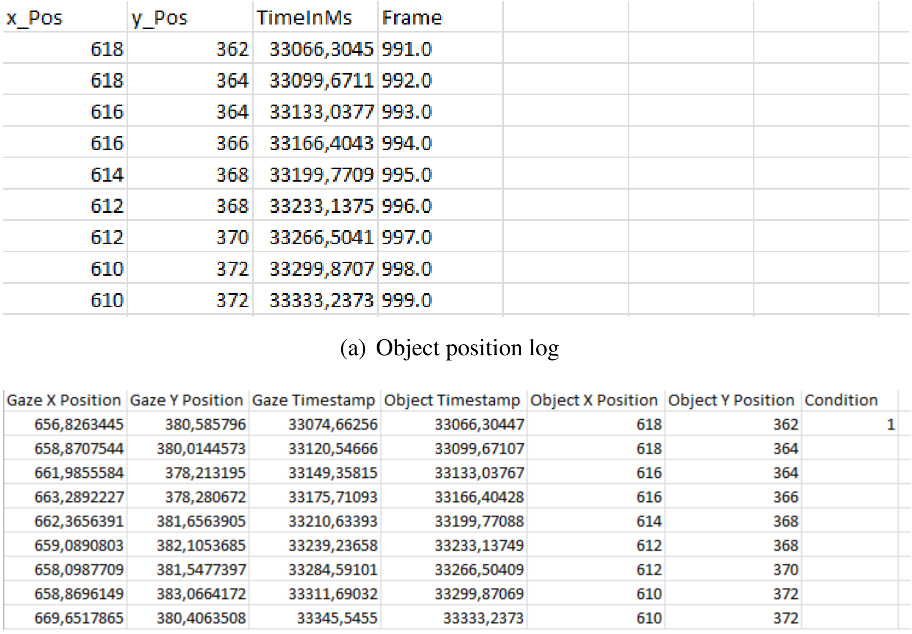
(B) False positive rate The number of times the correlation between the participant's gaze data and validation data exceeds a given threshold, calculated per participant in percentage

(C) Rate difference The difference between the true positive rate and the false positive rate (true positive - false positive), calculated per participant in percentage

(D) Time of first selection The index of the window that first exceeded the threshold, collected per participant and converted into milliseconds

True positives are the result of the correlation between the gaze data and the position of the real stimulus If the correlation is high the algorithm successfully detected a match of eye

[5] tobiigaming.com



x_Pos	y_Pos	TimeInMs	Frame
618	362	33066,3045	991.0
618	364	33099,6711	992.0
616	364	33133,0377	993.0
616	366	33166,4043	994.0
614	368	33199,7709	995.0
612	368	33233,1375	996.0
612	370	33266,5041	997.0
610	372	33299,8707	998.0
610	372	33333,2373	999.0

(a) Object position log

Gaze X Position	Gaze Y Position	Gaze Timestamp	Object Timestamp	Object X Position	Object Y Position	Condition
656,8263445	380,585796	33074,66256	33066,30447	618	362	1
658,8707544	380,0144573	33120,54666	33099,67107	618	364	
661,9855584	378,213195	33149,35815	33133,03767	616	364	
663,2892227	378,280672	33175,71093	33166,40428	616	366	
662,3656391	381,6563905	33210,63393	33199,77088	614	368	
659,0890803	382,1053685	33239,23658	33233,13749	612	368	
658,0987709	381,5477397	33284,59101	33266,50409	612	370	
658,8696149	383,0664172	33311,69032	33299,87069	610	372	
669,6517865	380,4063508	33345,5455	33333,2373	610	372	

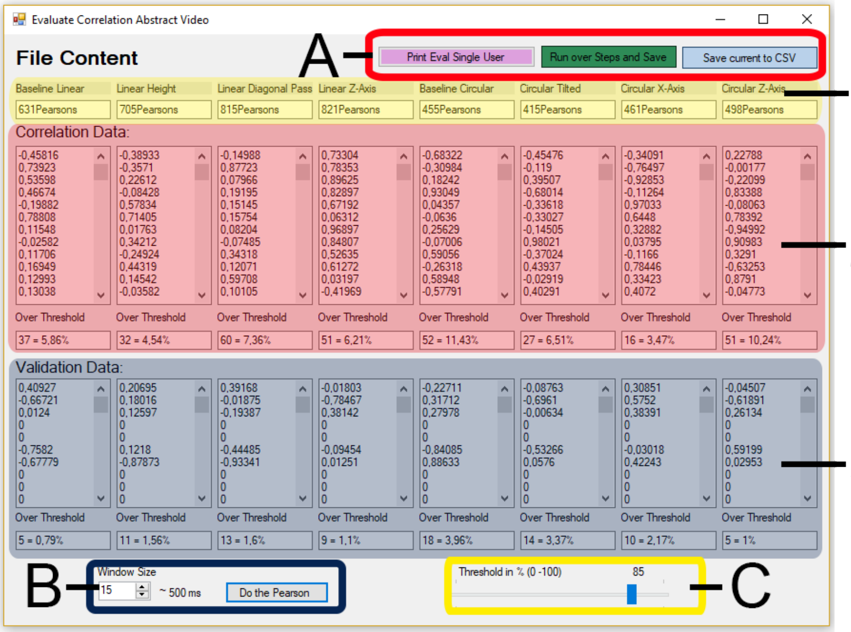
(b) Combined log

Figure 3.3: Pre-Study: An example of the log files created for the evaluation

and object movement. The number of times the correlation is exceeding a chosen threshold compared to how often the correlation is calculated is represented by the (A) true positive rate. It is calculated using the gaze data collected during the respective session and the position of the stimulus (red ball). To obtain this position an *object position log* needs to be created, by running the abstract video file (see fig. 2(c)) through an image processing script implemented with the Python library of OpenCV[6]. The output of that script is a *.csv-file listing the centre of the white circle in x- and y-coordinates ordered by their time stamps, additionally the file contains the index of the frame (see fig.3(a)). This file is subsequently combined with the gaze log file either produced by the Pupil software (real world environment) or the programmed video player (abstract and original video environment) by matching the gaze and object data over the nearest available time stamp. In cases were there was no movement condition executed, thus no motion of the stimulus, no log entry is created by the image processing script. This way during the matching of gaze and object data only periods are considered where both data is available. The resulting file (see fig. 3(b)) now shows the x- and y-coordinates of the participants gaze with the according time stamp and the matched object time stamp with the associated x-/y-coordinates. The movement condition has to be added manually (numbered 0-7) in order to provide that information to the automatic evaluation tool (see section 3.3.2 Evaluation Tool).

As the strength of the algorithm is not only determined by finding a high number of correct matches (true positives) but also by rejecting other object movements as a match, we introduce (B) the false positive rate. To get (B) we first calculate the correlation between the gaze data and a provided validation data set of a previous study. Thereby we get falsely positive matched movements in case the calculated correlation exceeds the threshold. The number of false positives compared to how often the correlation is calculated is the (B) false positive rate. Before we applied the data set of a previous study we tried several approaches to validate the Pursuits selection. First we correlated the participants data with other movement conditions of the same participant (e.g. used the gaze data of *Baseline Linear* and correlated it with object data from

[6] opencv.org/



File Content

Baseline Linear	Linear Height	Linear Diagonal Pass	Linear Z-Axis	Baseline Circular	Circular Tilted	Circular X-Axis	Circular Z-Axis
631Pearsons	705Pearsons	815Pearsons	821Pearsons	455Pearsons	415Pearsons	461Pearsons	498Pearsons

Correlation Data:

-0,45816	-0,38933	-0,14988	0,73304	-0,68322	-0,45476	-0,34091	0,22788
0,73923	-0,3571	0,87723	0,78353	-0,30984	-0,119	-0,76497	-0,00177
0,53598	0,22612	0,07966	0,89625	0,18242	0,39507	-0,92853	-0,22099
0,46674	-0,08428	0,19195	0,82897	0,93049	-0,68014	-0,11264	0,83388
-0,19882	0,57834	0,15145	0,67192	0,04357	-0,33618	0,97033	-0,08063
0,78808	0,71405	0,15754	0,06312	-0,0636	-0,33027	0,6448	0,78392
0,11548	0,01763	0,08204	0,96897	0,25629	-0,14505	0,32882	-0,94992
-0,02582	0,34212	-0,07485	0,84807	-0,07006	0,98021	0,03795	0,90983
0,11706	-0,24924	0,34318	0,52635	0,59056	-0,37024	-0,1166	0,3291
0,16949	0,44319	0,12071	0,61272	-0,26318	0,43937	0,78446	-0,63253
0,12993	0,14542	0,59708	0,03197	0,58948	-0,02919	0,33423	0,8791
0,13038	-0,03582	0,10105	-0,41969	-0,57791	0,40291	0,4072	-0,04773
Over Threshold	Over Threshold	Over Threshold	Over Threshold	Over Threshold	Over Threshold	Over Threshold	Over Threshold
37 = 5,86%	32 = 4,54%	60 = 7,36%	51 = 6,21%	52 = 11,43%	27 = 6,51%	16 = 3,47%	51 = 10,24%

Validation Data:

0,40927	0,20695	0,39168	-0,01803	-0,22711	-0,08763	0,30851	-0,04507
-0,66721	0,18016	-0,01875	-0,78467	0,31712	-0,6961	0,5752	-0,61891
0,0124	0,12597	-0,19387	0,38142	0,27978	-0,00634	0,38391	0,26134
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
-0,7582	0,1218	-0,44485	-0,09454	-0,84085	-0,53266	-0,03018	0,59199
-0,67779	-0,87873	-0,93341	0,01251	0,88633	0,0576	0,42243	0,02953
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
Over Threshold	Over Threshold	Over Threshold	Over Threshold	Over Threshold	Over Threshold	Over Threshold	Over Threshold
5 = 0,79%	11 = 1,56%	13 = 1,6%	9 = 1,1%	18 = 3,96%	14 = 3,37%	10 = 2,17%	5 = 1%

Window Size 15 ~ 500 ms

Threshold in % (0 -100) 85

Figure 3.4: The evaluation tool with three information zones (1-3) and three interaction zones (A-C)

C-Tilted). This however resulted in an equally high false and true positive rates. Presumably because the eight movements have a similar trajectory regarding the on-screen coordinates, as they were not designed to be very different in the beginning. Furthermore we used the participants data and reversed the respective object data as well as combinations of both mentioned approaches. Lastly artificially generated data was used such as sinus waves or linear functions also resulting in high false positive rates. Therefore a real object movement data set from a previous study was chosen that subsequently yielded lower false positive rates.

The best possible case for the produced outcome of the algorithm is a high true positive rate and a low false positive rate, that is why we also evaluate the (C) rate difference. Lastly we investigate how long it took until the first true positive was calculated with (D) Time of the first selection. Therefore we simply collect the index of the time window whose data produced the first true positive. For the evaluation process this index is multiplied by the window size in order to get the selection time in milliseconds.

3.3.2 Evaluation Tool

After realising that a manual evaluation using Microsoft Excel[7] would be inconclusive as it had only enabled us to compare single participants, we designed an evaluation tool. It is implemented with C# on the .NET Framework 4.5.2 and takes the combined log file (see fig. 3(b)) as input. The final version was used to generate the pre-study evaluation data as well as the usability study data (see section 4 Developing and Testing Use Cases). Prior to the version explained in this section, implementations to only evaluate the single participants, one without the *validation data* and one with less output functions were developed.

The tool can be separated in three information zones (see fig. 3.4, coloured boxes: 1-3) and three interaction zones (see fig. 3.4, framed boxes: A-C). Zone 1 displays the eight names of the movement conditions with the respective number of overall correlations performed for this

[7] products.office.com/de-de/excel

condition (changes depending on window size). Zone 2 was named *correlation data* in the process before above terminology was introduced. It displays data regarding the true positive values: the upper row of boxes lists the results of all executed correlations, the bottom row displays the true positive working rate (see below). Zone 3 is dedicated to the false positive calculations (called *validation data* at the time). Analogous to zone 2, the upper row lists all results of the correlation with the validation data set and the bottom row displays the false positive working rate.

The interaction zones are used to either manipulate the parameters for the Pursuits algorithm or generate output files. In region B the window size in lines is set. As we worked based on 30fps video files we assumed that for most calculation the difference between time stamps: $(timeArrayEntryX + windowsize) - timeArrayEntryX$ would result in $windowsize * 33$ milliseconds. After each change of the window size parameter the *Do the Pearson*-button has to be pressed to execute the algorithm with the new parameter, hereby changing the number of calculations presented in zone 1. Zone C lets you change the threshold which needs to be exceeded in order to register as a movement match/true positive. The displays in 2 and 3 update automatically when using the slider.

The buttons arranged in zone A print the current displayed working rate information into an *.csv-file (*'Save current to CSV'*, right) or generate a *.csv-file containing the working rates for the given window size over several thresholds ($0.1 - 1.0$ in 0.5 steps) (*'Run over Steps and Save'*, middle). The last button *'Print Eval Single User'* (left), generates a *.csv-file containing all evaluation aspects (true positive rate, false positive rate, rate difference, time of first selection) listed for each participant individually calculated for the same range of thresholds as mentioned above, which are the base for the evaluation results in section 3.4 Pre-Study: Results. Because gaze and object data are not separated by participant for the above mentioned tool-display, it only provides the working rate of the algorithm (which will not be used in the following evaluation). In order to get the values divided by participant the *'Print Eval Single User'* has to be pressed.

3.3.3 Evaluation Methodology

The evaluation was carried out using the data of all 24 participants of the pre-study. In order to find out how rates differ regarding to the window size we decided to compare three sizes that have been tested for on-screen applications. As the smallest window we chose 500ms as used by Vidal et al. [45]. As they concluded that Pursuits performs better for larger windows we also applied 1000ms [10] and 2000ms [26]. Even though a response time of at least two seconds might be impractical for applications we wanted to compare the performance of Pursuits with real world targets on larger time windows.

We investigated all four evaluation aspects regarding those three time windows for all three environments. Firstly we compared the results of the individual environmental conditions regarding the three different window sizes, aiming at finding an ideal window size and threshold in order to be able to compare the environments to each other. An ideal window size would present (significantly) better results for all evaluation aspects. An ideal threshold would constitute by being the first value to present an (significantly) higher rate than the ones before or signalling a saturation effect regarding the selection (e.g. differences in the true positive rate are no longer significant).

Next we used the evaluation aspects to compare the performance of Pursuits in regards to the three different environments. The results of this evaluation are discussed in section 3.4 Pre-Study: Results. Tests for statistical significance were conducted with a two-way repeated measures ANOVA.

3.4 Pre-Study: Results

In the following section the results of the pre-study are discussed. Beginning with an analysis of the three window sizes (500ms, 1000ms and 2000ms) for each environmental condition (real world, original video, abstract video). Followed by an comparison of the environments.

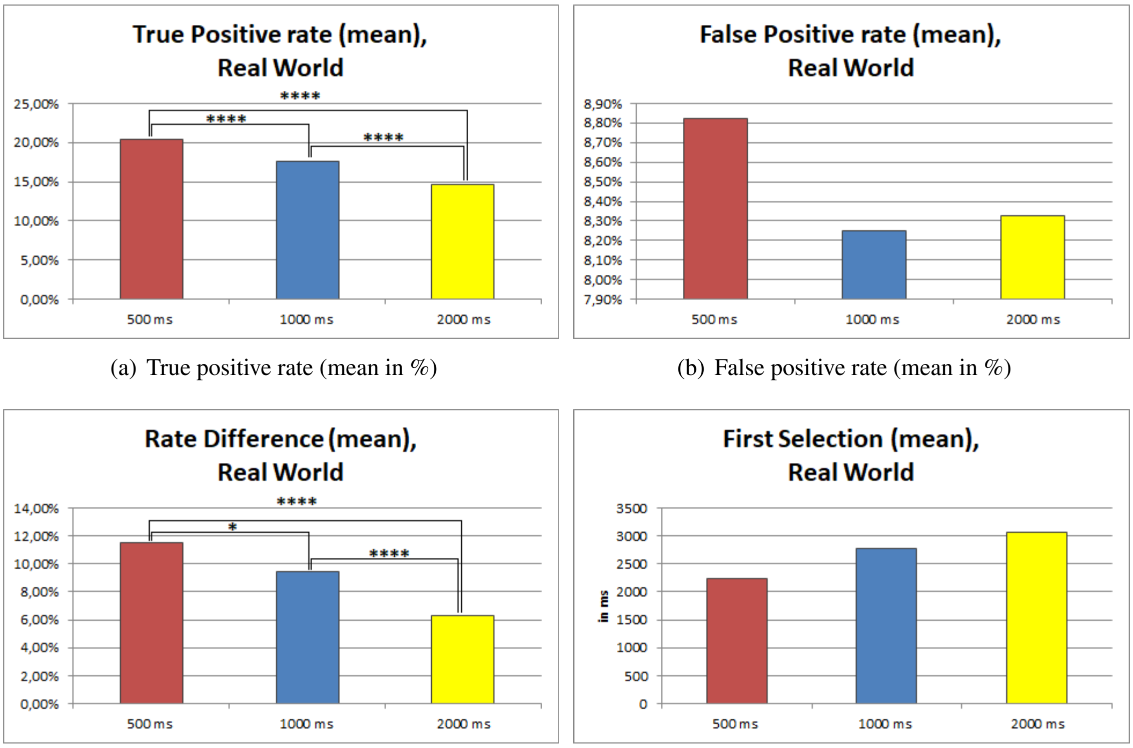
3.4.1 Comparing Window Sizes

Pursuits was applied with a 500ms, 1000ms and 2000ms time window and results of the individual participants printed for thresholds between 0.1 and 1.0 in 0.5 steps. In order to get an idea of the performance with these window sizes we first compare the results within the environment conditions. For understanding the impact movement types have, we distinguish not only between the eight movement conditions but also between movements without depth factor (grouped results from the two baseline conditions: Baseline Linear and Baseline Circular) and with depth factor (grouped results of the alternation: C-Tilted/-xAxis/-yAxis and L-Height/-DiagonalPass/-zAxis) referred to as dimensions in the following. Most interesting for us is whether there are differences in results for movement that had a depth factor or not, when they were first carried out in the laboratory setting of the first session.

Real World Environment We analysed the data from *real world environment* regarding the *true positive rate*, *the false positive rate*, *the rate difference* and the *time of the first selection* (see section 3.3.1 Evaluation Aspects) over the different window sizes. Starting with the *true positive rate* (see fig. 3.5 (a)), which presents to be significantly different comparing the three time windows (ANOVA:$p < .0001$, Greenhouse-Geisser adjustment was used to correct violations of sphericity). A Bonferroni corrected post-hoc test reveals statistically significant differences between the rate for all three window sizes. Using 500ms resulted in the best true positive rate (mean: 20.4% SD: 1.6%), being significantly better than the rate for 1000ms (mean: 17.7% SD: 1.8%, $p < .0001$) and 2000ms (mean: 14.6% SD: 1.9%, $p < .0001$). The difference between the true positive rate for 1000ms and 2000ms (mean difference: 3.1 percentage points) is also significant ($p < .0001$). Analysis pointed out that there is a significant difference between true positive rates for the different movement conditions (ANOVA: $p < 0.0001$. Greenhouse-Geisser corrected), however none between movements with and without a depth aspect in general. There are no significant differences between false positive rates (see fig. 3.5 (b), with almost equally low false positives rates Pursuits rejects false matches with a 1000ms window (mean: 8.25% SD: 0.4%) as good as with a 2000ms window (mean: 8.33% SD: 0.6%). And thereby better than with a 500ms window, where false positive rates are slightly higher (mean: 8.8% SD: 0.4%).

The significant differences for true positive rates translate onto the rate difference (see fig. 3.5 (c)) presenting significant differences between window sizes (ANOVA: $p < .0001$). Bonferroni corrected post-hoc tests expose that the rate difference produced with a 500ms-Pursuits (mean: 11.5 percentage points SD: 1.4 percentage points) is significantly bigger than the rate difference from the 1000ms-Pursuits (mean: 9.4 percentage points SD: 1.7 percentage points, $p = .013$) as well as from the 2000ms-Pursuits (mean: 6.3 percentage points SD: 1.8 percentage points, $p < .0001$). Again differences between the rate difference for single movement conditions are significant, however not between with or without depth aspect.

Due to too many missing cases (no selection made at all for higher thresholds) the two-way repeated measures ANOVA could not be performed for the time of the first selection, therefore descriptive statistics are used presenting the time of first selection. With an average time of 2764.34ms until the first selection the 500ms window for Pursuits responses the quickest to selections. The minimal time need to register a true positive is one time window (= 500ms) and



(a) True positive rate (mean in %)

(b) False positive rate (mean in %)

(c) Rate difference (mean in percentage points displayed with %)

(d) Time of first selection (mean in ms)

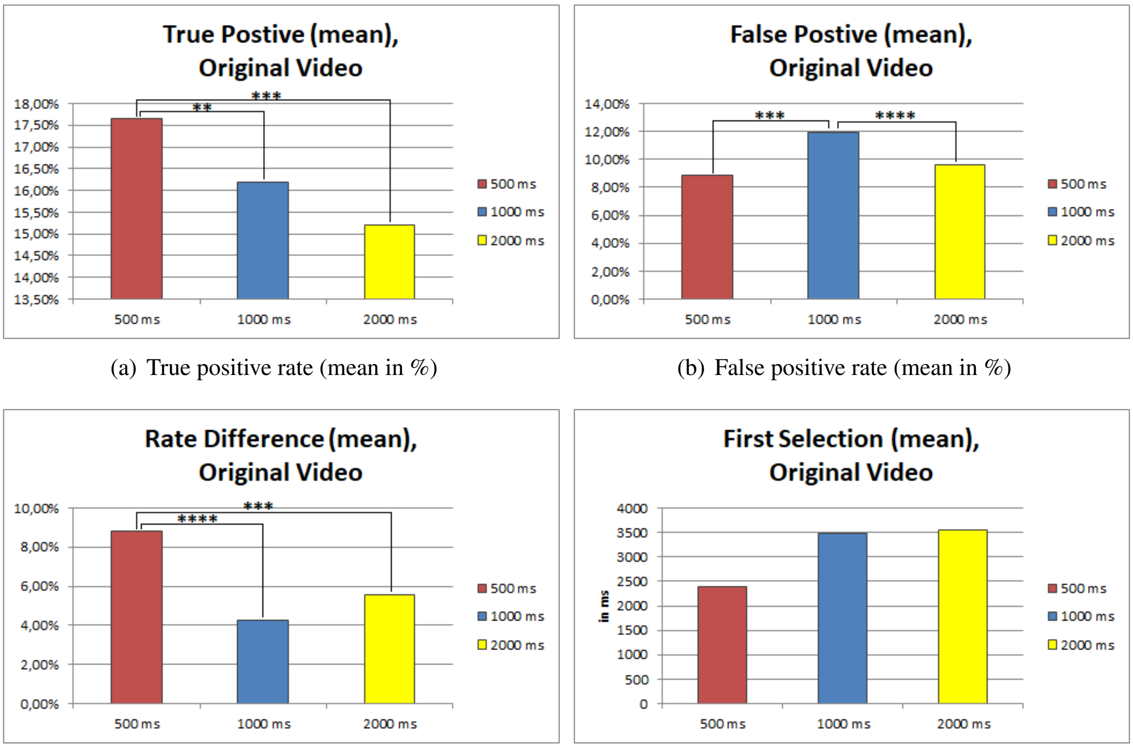
Figure 3.5: Real world environment: Results evaluation aspects for 500/1000/2000ms
The use of a 500ms window outperforms 1000/2000ms in three of four evaluation aspects:
true positive rate (significant), rate difference (significant) and selection time.

the longest a participant would have waited for selection is 16 seconds. Logically the earliest first selection with a larger time window can only take place later, so for 1000ms and respectively 2000 my the earliest selection is registered after 1000ms and 2000ms (one time window in each case). On average a participant has to wait 5107.02ms with a 1000ms window and 5907,03ms with a 2000ms window. The latest selection with a 1000ms window is registered after 19 seconds and after 20 seconds for 2000ms (see fig. 3.5 (d)).

Analyses regarding an ideal threshold turn out to be conclusive. All threshold-steps produce significant results to one another (e.g. the number of true positives yielded with a 0.60 is significantly lower than 0.65) or none at all.

Despite producing the highest false positive rate the 500ms returns overall better results than the other time windows, by having an higher true positive rate, even high enough to also have the largest rate difference and selecting true positives the quickest.

Original Video Environment Secondly we analysed the data from *original video* regarding the *true positive rate*, *the false positive rate*, *the rate difference* and the *time of the first selection* (see section 3.3.1 Evaluation Aspects) over the different window sizes. ANOVA reveals significant differences between true positive rates (see fig. 3.6 (a)) of several window sizes ($p < .0001$). Pairwise comparison of a Bonferroni corrected post-hoc test exposes true positive rates of Pursuits with 500ms (mean: 17.6% SD: 0.8%) are significantly higher than 1000ms (mean: 16.2% SD: 0.8%, $p = .001$) as well as 2000ms (mean: 15,2% SD: 1.0%, $p = .0003$). The algorithm with



(a) True positive rate (mean in %)

(b) False positive rate (mean in %)

(c) Rate difference (mean in percentage points displayed with %)

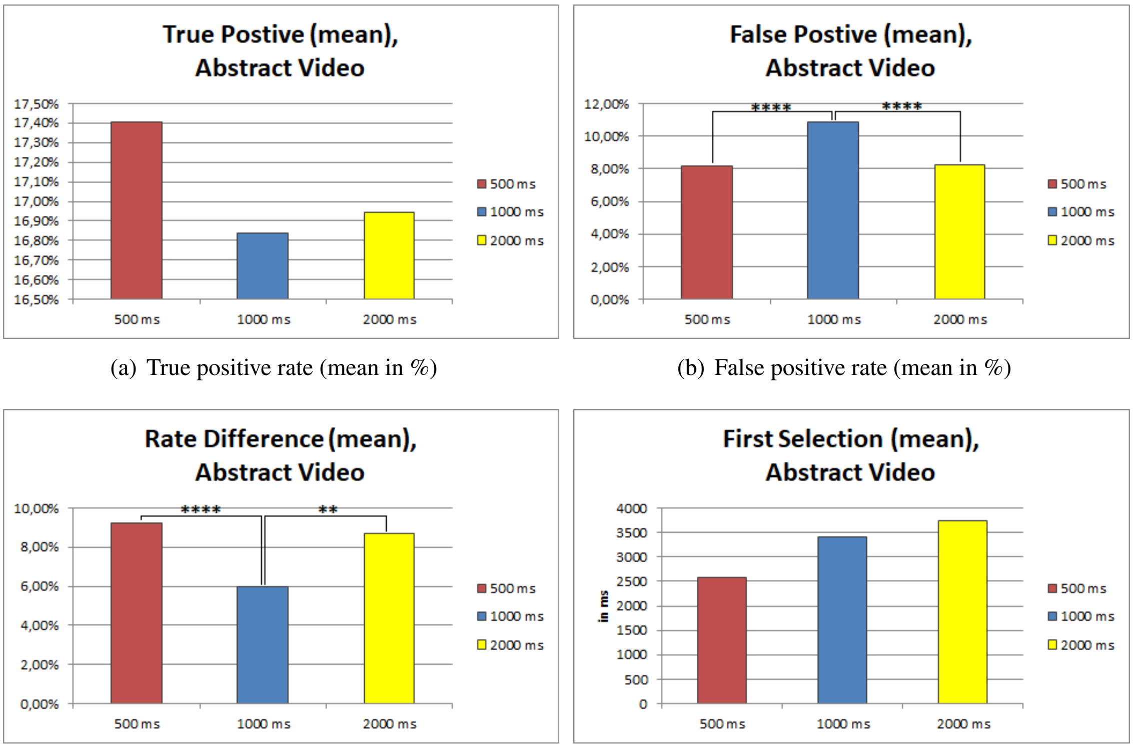
(d) Time of first selection (mean in ms)

Figure 3.6: Original Video environment: Results evaluation aspects for 500/1000/2000ms
The use of a 500ms window outperforms 1000/2000ms in all four evaluation aspects:
true positive rate, false positive rate, rate difference (all significant) and selection time.

a 2000ms window also performs worse than with a 1000ms window, this difference is not significant though.
The Pursuits in the *original video* environment also resulted in statistically significant different true positive rates for single movement types not, however, between movements with or without depth factor in general.

Significant differences can also be found for the false positive rate (see fig. 3.6 (b)) with a $p < .0001$ calculated with ANOVA. Post-hoc Bonferroni shows false positive rates are significantly higher for 1000ms (mean: 11.9% SD: 0.6%) than for 500ms (mean: 8.9% SD: 0.4%, $p = .0004$) and for 2000ms (mean: 9.6% SD: 0.7%, $p < .0001$). Performance differences for false positive rates with a 500ms and a 2000ms window are negligible.

There are significant differences in rate differences on *original video*, violations of sphericity were corrected using Greenhouse-Geisser adjustments producing a $p < .0001$. Bonferroni corrected post-hoc tests shows that the rate difference is significantly higher when a 500ms window is used (mean: 8.8 percentage points SD: 0.9 percentage points). It provides a twice as big rate difference as 1000ms-Pursuits (mean: 4.5 percentage points SD: 1.1 percentage points, $p < .0001$) and also a larger rate difference than produced with 2000ms (mean: 5.6 percentage points SD: 1.3 percentage points, $p = .0002$). Further difference do not come up as significant. Again the different movement conditions yield significantly different rate differences, these do not appear to be between movements with and without depth in general.



(a) True positive rate (mean in %)

(b) False positive rate (mean in %)

(c) Rate difference (mean in percentage points displayed with %)

(d) Time of first selection (mean in ms)

Figure 3.7: Abstract Video environment: Results evaluation aspects for 500/1000/2000ms
The use of a 500ms window outperforms 1000/2000ms in all four evaluation aspects:
true positive rate, false positive rate (significant), rate difference and selection time.

Due to too many missing cases (no selection made at all for higher thresholds) the two-way repeated measures ANOVA could not be performed for the time of the first selection, therefore descriptive statistics are used presenting the time of first selection. With a 500ms time window first true positives are detected the quickest. With an average true positive match after 2393.06ms a system response can be initiated a second earlier than when using a 1000ms window (mean time of first selection: 3486.76ms) or a 2000ms (mean time of first selection: 3544.78ms). For all three cases the minimal time need to register a match was one time window. Other than the average selection time the longest time periods are relatively equal with 18 seconds for the shortest time window and 20 seconds each for the two longer ones.

Analyses regarding an ideal threshold turn out to be conclusive. All threshold-steps produce significant results to one another (e.g. the number of true positives yielded with a 0.60 is significantly lower than 0.65) or none at all.

For the use with an *original video* environment Pursuits with a 500ms window produced the best results in all tested categories. Under these settings selection of the most true positives happens the quickest while also rarely registering false positive matches resulting in the highest rate difference.

Abstract Video Environment In the *abstract video* condition no significant differences for the true positive rates can be reported. Pursuits produced equal true positive rates over all time windows of approximately 17% (see fig. 3.7 (a)). However Pursuits performs differently for single movement conditions, not for movements with or without depth in general though. Nevertheless some window sizes produced such low false positive rates, that for rate difference as well as false positive significant difference are shown.

ANOVA revealed significant differences in the false positive rates overall ($p < .0001$) and the Bonferroni corrected post-hoc test reveals that Pursuits executed with a 1000ms window (mean: 10.9% SD: 0.4%) produces significantly higher false positive rates that using 500ms (mean: 8.2% SD: 0.4%, $p < .0001$) and 2000ms (mean: 8.2% SD: 0.5%, $p < .0001$). With an equal false positive rate using 500ms or 2000ms register least of all false positives.

Due to the varying low false positive rates significant differences in the amount the rates differ are apparent ($p < .0001$, Greenhouse-Geisser adjustments were used to correct violations of sphericity). Post-hoc Bonferroni shows that with a 1000ms window Pursuits results in a significantly smaller difference between true and false positive rates (mean: 6.0 percentage points SD: 1.2 percentage points) comparing to the use of a 500ms window (mean: 9.2 percentage points SD: 1.1 percentage points, $p = .002$) and 2000ms (mean: 8.7 percentage points SD: 1.9 percentage points, $p = .009$). With a mean difference of 0.5 percentage points, under 500ms a irrelevantly higher difference than under 2000ms was archived. For false positive rates and rate difference ANOVA also yields statistically significant results regarding single movement types, however non between dimensions.

Due to too many missing cases (no selection made at all for higher thresholds) the two-way repeated measures ANOVA could not be performed for the time of the first selection, therefore descriptive statistics are used presenting the time of first selection. The quickest a selection is made is one time window for all three time windows (so 500ms, 1000ms and 2000ms) the longest a user would have waited for a response is 18 seconds (size 500ms), 20 seconds (2000ms) and 21 seconds (1000ms). Also on average using a 500ms window results in quickest selections (2582.35ms), almost one second before Pursuits with the 1000ms time window would react (3400,51ms) and equally quicker under the 2000ms window (3727.52ms).

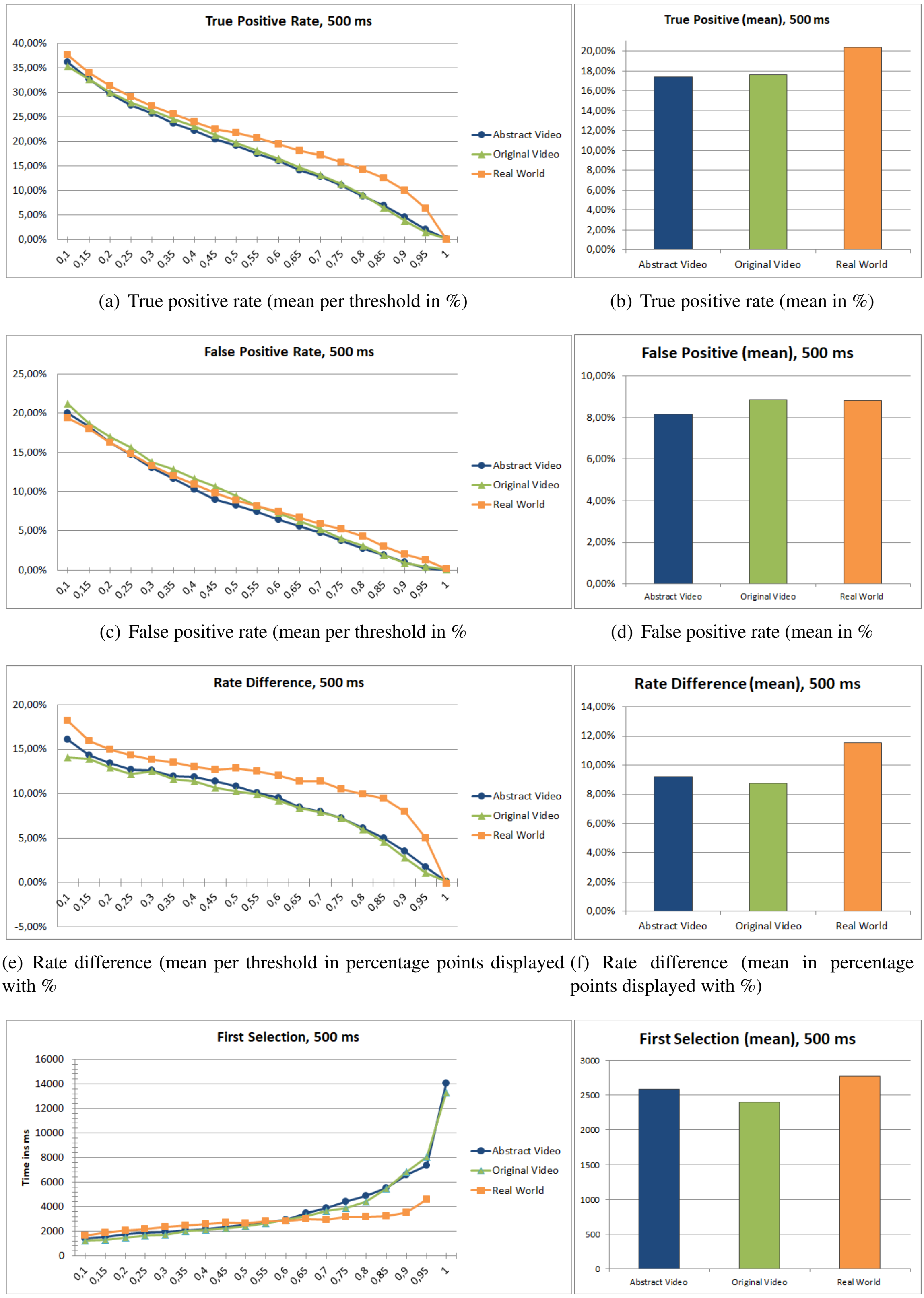
Analyses regarding a ideal threshold turn out to be conclusive. All threshold-steps produce significant results to one another (e.g. the number of true positives yielded with a 0.60 is significantly lower than 0.65) or none at all.

Despite no window size produces better true positive rates on a *abstract video* environment than the others, Pursuits all other evaluation aspects hint at using 500ms, as the biggest rate difference because of the low false selection rate is archived and response time is the shortest.

3.4.2 Comparing Environments

The results of the evaluation above suggests that shorter time spans, in our case 500ms, lead to a better Pursuits performance for real world stimuli. Therefore we compare the evaluation aspects *true positive rate*, *the false positive rate*, *the rate difference* and the *time of the first selection* (see 3.3.1 Evaluation Aspects) of the different environment conditions using the Pursuits results with a 500ms window. Not only are we interested in difference between environments but also the influence the dimensions (with or without depth factor) combined with the environment has on the single evaluation aspects.

True Positive Rates As can be seen in fig. 3.8 (a)+(b) there are interesting differences in the true positive rates for real world, original video and abstract video. Especially for higher thresholds there are better true positive rates in the *real world* environment (mean: 20.4% SD: 1.6%). How-



(a) True positive rate (mean per threshold in %)

(b) True positive rate (mean in %)

(c) False positive rate (mean per threshold in %

(d) False positive rate (mean in %

(e) Rate difference (mean per threshold in percentage points displayed with %

(f) Rate difference (mean in percentage points displayed with %)

(g) Time of first selection (mean per threshold in ms)

(h) Time of first selection (mean in ms)

Figure 3.8: Comparing environments real world, original video and abstract video.
Using window size = 500ms
Despite interesting differences between real world and video environments, no significance can be reported.

ever the differences to the selection rate of *original video* (mean: 17.6% SD: 0.8%) or *abstract video* (mean: 17.4% SD: 1.0%) turn out not to be statistically significant. Taking a closer look into the performance for the combined factor *environment* $*$ *dimension* it stands out that the mean true positive rate in the *real world* environment is minimal better for movements with depth factor (mean: 20.5% SD: 1.6%) than for movements without depth factor (mean: 20.0% SD: 2.2%). Other than for *original video* (with: mean: 17.6% SD: 0.8%; without: mean: 17.7% SD: 1.4%) or *abstract video* (with: mean: 16.9% SD: 1.1%; without: mean: 19.0% SD: 1.2%) where the true positive rate decreases for stimulus movements into the depth of the three dimensional space.

False Positive Rates The false positive rates (see fig. 3.8(c)+(d)) are relatively equal in all environments, which is confirmed by the two-way repeated measures ANOVA, that does not yield significant differences. Using Pursuits on an abstract environment produced slightly less false positives (mean: 8.2% SD: 0.4%) than in a real world environment (mean: 8.8% SD: 0.4%) or on an original video environment (mean: 8.9% SD: 0.4%). Although the same way as for the true positive rate, false positives rate differ minimally for *environment* $*$ *dimension*. With Pursuits in the real world environment also producing higher false positive rates, this being a drawback in this case, for movements with depth aspect (with: mean: 8.9% SD:0.5%; without: mean: 8.5% SD: 0.5%). *Original video* as well as *abstract video* as an environment lead to lower false positive rates for movements that had a depth factor when being performed (original video: mean: 8.8% SD: 0.4%; abstract video: mean: 7.9% SD: 0.4%) than without a depth factor (original video: mean: 9.0% SD: 0.8%; abstract video: mean: 9.1% SD: 0.6%).

Rate Difference Combining the knowledge from the analyses above it is already obvious that ANOVA does not show any significant difference for the rate difference (see fig. 3.8(e)+(f)) regarding the three environmental conditions. Because of the higher true positive rate the rate difference for *real world* is bigger (mean: 11.5 percentage points SD: 1.4 percentage points) than for *abstract video* (mean: 9.2 percentage points SD: 1.1 percentage points) and *original video* (mean: 8.8 percentage points SD: 0.9 percentage points). This trend is also visible when comparing non-depth conditions to depth conditions, although the divergences for each environmental condition is even smaller than for the evaluation aspects above (real world without: 11.5 percentage points SD:2.1 percentage points; with: 11.6 percentage points SD: 1.2 percentage points // original video without: 8.7 percentage points SD: 1.8 percentage points; with: 8.8 percentage points SD: 1.0 percentage points // abstract video without: 9.8 percentage points SD: 1.3 percentage points; with: 9.0 percentage points SD: 1.3 percentage points)

Time of first selection Figure 3.8(g) shows that the time of the first selection in the *real world* environment stays relatively constant even for higher threshold compared to the two video environments where selection time increases for higher thresholds. However figure 3.8(h) reveals that the overall average selection time is higher for the *real world* environment (mean: 2764.34ms) than for the *abstract video* environment (mean: 2582.35ms) and the *original video* environment (mean: 2393.06ms). Due to too many missing cases (no selection made at all for higher thresholds) the two-way repeated measures ANOVA to test statistical significance of these observations could not be performed for the time of the first selection.

3.5 Lessons learnt

The pre-study's results show that applying Pursuits to gaze and object data from the *real world* works in general and does not lead to worse results than on-screen applications. The evaluation shows that real world scenarios require shorter time windows for Pursuits to perform best no matter what environment is used. In our case outcomes for 500ms exceeded on all environments for almost every evaluation aspect. Using the smallest time window outperforms on three of the four tested aspects for all environments producing slightly higher false positive rates. In order to learn whether Pursuits tests for real world scenarios can be conducted with video prototypes, we compared the outcome of different environments. However no statistically significant divergence in performance between environments for the use of real world material can be reported. We did not conduct any interviews or surveys after the sessions since we were primarily interested in collection data. Still some users commented after the last environmental condition, that watching the video was either much more demanding, especially for the *original video*(stated by participant 'Berlin'), or distracting (stated by participant 'Oslo'). Opinions to why it is distracting varied. Participant 'Oslo' stated: 'the scene is simply overwhelming and it is surprisingly harder to focus on the ball even though I did it in the real world'. And Participant 'Cardiff' pointed out technical reasons such as 'I sometime switched focus and looked at the reflections on the screen instead of the target'. Due to these comments and the divinations of the *real world* environment rates, in contrast to the almost equal performances on the two video environments, we conclude that it is not sufficient to test real world scenarios on screen as the same results cannot be expected, concerning Pursuits performance as well as user assessment.

Main findings in short:

- Pursuits scenarios in the real world require shorter time windows.
- Our smallest window of 500ms significantly outperformed other times on almost all aspects
- No statistically significant for the comparison of environments, despite slight deviations
- Users state watching videos comes with distractions
- Therefore we conclude that it is not sufficient to test real world scenarios on screen
- In doing so the same results cannot be expected

4 Developing and Testing Use Cases

Objects that eyes pursue cannot only be displayed on a screen, but are actually all around us. A bird flying by, a car passing or a person walking by . Even stationary objects like buildings can appear to be moving, if the user is moving. Driving a car through a landscape and the monuments move past the window or taking the usual route through the park when jogging and the benches and bins start coming towards you. The results of a brainstorming session to identify suchlike use cases and their underlying evaluation aspects are discussed in the first parts of the following chapter. Eventually the user study of two chosen scenarios and its results are reviewed.

4.1 Brainstorming Session

We organised a brainstorming session to identify use cases for Pursuits in the real world. Therefore the participants were introduced to the concept of Pursuits on-screen and the requirements for it to work (e.g. moving stimulus that the eye can pursue). We asked to come up with fitting scenarios where if possible also the natural gaze behaviour is exploited. While being told to ignore potential technical limitations of eye tracking or getting information about the environment. The seven participants came up with a number of ideas (see Fig. 4.1), which are described in detail in this section.

Information The main idea behind the category 'Information' is to display information about things that are moving either because the object itself is in motion or the user is moving past the object. In the *Zoo* scenario the animals would present a moving target. By following one specific animal the observer receives information about either that individual or the species overall. The information can either be sent to a smart phone or displayed on a screen at the compound. A similar approach is taken for the use cases *sport events* and *running sushi*. The user often naturally follows the object of interest, in these cases the animal at the zoo, the football player at a game or the desired plate at the restaurant, which can be used as target when enquiring information. A slightly differing concept would underlie *movies*, where Pursuits is again used on targets provided on-screen but this time not deliberately to interact. The user watches a film and wants know more about an actors work, by pursuing them the relevant information can be transferred to e.g. the smart phone (for a non-disruptive viewing pleasure) or displayed directly on-screen.

While these cases are about providing information in a entertainment sense the scenarios *airport/air plains*, *assembly line* and *surveillance* cover a more professional aspect for staff members in their work environment. For instance employees working at an airport tower pursuing landing or departing air planes, can get information about them in order to coordinate them. Then even be able to directly initiate communication with the plane's crew. With the targets moving more freely, this case relates to e.g. the *zoo* example form above, whereas products on an assembly line are more similar to dishes on a line at a restaurant. Both would work with the same concept, just in the professional environment the observer is given information about e.g. time of production or destination. Disposing of faulty products on sight is also a possibility with Pursuits. For people sitting at surveillance desks using Pursuits can offer an interaction method of the camera controls, based on the same concept as the *movie* situation, the user would follow their person of interest and the next available camera setting is automatically chosen.

All the situations until now assume a moving target and a stationary observer, but it is also possible to offer or communicate information about stationary object or moving objects from a observer perspective as they are in motion. Such as in the scenarios *supermarket* and *behaviour prediction*. So is information about a supermarkets range of a product displayed at the end of their trolley when the costumer inspects a section while passing by. Furthermore in combination with

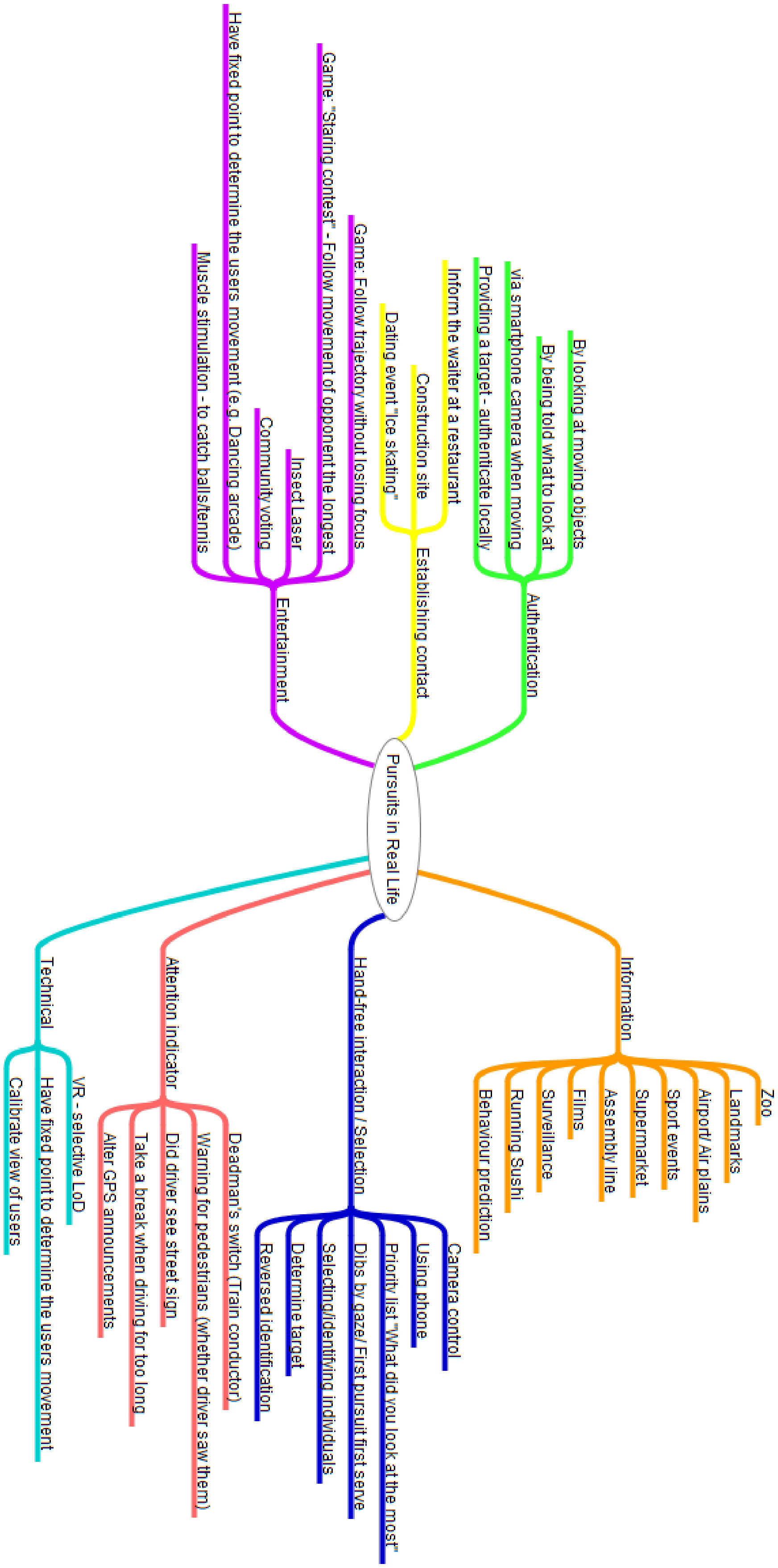

Pursuits in Real Life
Entertainment
Game: Follow trajectory without losing focus
Game: "Staring contest" - Follow movement of opponent the longest
Insect Laser
Community voting
Have fixed point to determine the users movement (e.g. Dancing arcade)
Muscle stimulation - to catch balls/tennis
Establishing contact
Inform the waiter at a restaurant
Construction site
Dating event "Ice skating"
Authentication
By looking at moving objects
By being told what to look at
via smartphone camera when moving
Providing a target - authenticate locally
Information
Zoo
Landmarks
Airport/ Air plains
Sport events
Supermarket
Assembly line
Films
Surveillance
Running Sushi
Behaviour prediction
Hand-free interaction / Selection
Camera control
Using phone
Priority list "What did you look at the most"
Dibs by gaze/ First pursuit first serve
Selecting/identifing individuals
Determine target
Reversed identification
Attention indicator
Deadman's switch (Train conductor)
Warning for pedestrians (whether driver saw them)
Did driver see street sign
Take a break when driving for too long
Alter GPS announcements
Technical
VR - selective LoD
Have fixed point to determine the users movement
Calibrate view of users

Figure 4.1: Categorised results of the use case brainstorming session

behavioural models predictions about objects of interest are made when the user pursuing it with their eyes, e.g. because a person is walking on the right side of the pavement they are more likely to turn right at the next possibility.

Hands-free Interaction/Selection Often gaze interaction is applied when hands-free interaction is desired for example to overcome distance or simply full hands. Selecting an individual object is easy when it is within arm's reach but gets difficult if it is further away as pointing becomes ambiguous. Determining and fixing a flying target (e.g. drone, plane (see *Information*), insects) with the eyes seems, according to the focus group, a relatively natural behaviour already. That gaze selection can be faster than mouse selection on a screen environment was found by Sibert and Jacob [37]. The next step would now be to also enable selection over gaze on real world moving targets, which can be archived using Pursuits. For objects like planes interaction ends with displaying information after making a selection, with the use case *determine target* interaction is taken a step further: for objects like drones, if one is the owner, the selection of one device can be part of a game or enable controlling mechanism for that gadget. In the military section Pursuits can be used to support aiming at moving targets.

The scenario *Selecting/identifying target* is also closely related to the *Information* category above but it grasps the more general concept of selecting an individual over greater distance or from a group, to start any kind of interaction. Besides information displaying the identified object can be selected for a special treatment e.g. a farmer can just by looking at it select an individual of their live stock and for example put it on the list for milking. A combination of this idea and *surveillance* (see *Information*) is presented with *reverse identification*, therefore the participants assumed that people unconsciously react to seeing a person of interest by pursuing them visually: Here a suspect is presented with several people, one of them being a victim. The participants guessed that it is possible the suspect will show more interest in their victim than in the random people. Based on the concept of selecting objects far away or out of reach for other reasons it *dibs by gaze/first pursue first serve*, which is thought to work as a reserving mechanism e.g. at a Sushi place or the subway, where you can "call dips" on a plate or a seat by looking at it.

Aiming more at alternative interaction methods when objects are within reach but direct interaction requires more effort e.g. when carrying bags in both hands or while running, is the use case *using phone*. The example during the focus group was directed at the use of phones but can be adapted to any device. It assumes that users are already connected to their phone (it was suggested that "a lot of people listen to music on their phone and/or use the earphones to talk on the phone anyway"). Imagine going home from shopping, bags in both hands and earphones connected to the phone, which is receiving a call. Instead of now having to put down the bags and search for the phone, a object in the environment is determine to function as the accept button another on as decline. By following one or the other the respective action will be triggered. The same mechanism is applied when the user is running and stopping to get to the phone is inconvenient.

The scenario named *camera control* again speaks more to a professional user. Pulling focus and keeping the person of interest in frame is a especially demanding for moving objects. With Pursuits both tasks can be achieved by a single camera operator. By following the person or object of interest (e.g. a new character coming into scene or a car at a race) in the real world the operator sets a target on which the camera focuses, additionally artistic framing aspects can be set on a preview screen and tracking with those parameters activated. Different variations were suggested, e.g. the following mechanism, once started, is always active leaving the operator to follow the object for the desired period opposed to selecting once and registering a object as the target until replaced by another one or tracking is turned of.

The above scenarios expect moving objects and a stationary user, the last idea *priority list* utilises the user's motion when walking through an exhibit to enable Pursuits. Instead of having to take note of interesting merchandise on-site, the user later receives an automatically generated weighted list. It presents articles they looked at the longest or most often. Essentially providing a priority list of the offered goods. Again the specific scenario made up during the session can be generalized.

Attention Indicator This category covers cases where the focus group's participants use gaze behaviour to indicate attention. Or in a way measure attention in traffic, where Pursuits can also help to suggest or predict behaviour of other traffic participants. It can also support communication between road users (*warning pedestrians*, *did driver see street sign*): sometimes a pedestrian is not sure whether the driver of a car has actually seen them approaching the street or is randomly slowing down, especially during the night when it is too dark to see the driver and make eye contact. If the driver has been pursuing the pedestrian walking for a while this can be an indicator that their slowing down is because of the pedestrian (should never be taken for certain though). Also communication between drivers can be supported in the way that a turn-intention is communicated to others, when the driver looked at several traffic signs implying so, in case the user does not indicate. Additionally drivers are supported with information based on pursued and read street signs. So can a warning be shown when a driver did not look at a GIVE WAY-sign. Or navigation system announcements can be altered based on behaviour of the driver in combination with their gaze behaviour (e.g. a set indicator in combination with the user following a recurring sign over a longer period leads to a dismissal of the next turn announcements, *alter GPS announcements*).

Another aspect is to see whether someone is alert and also able to act as for example necessary when conducting trains or driving for a long time. For train conductors gaze interaction with Pursuits replaces the physical *deadman's switch*, that has to be pressed every once in a while to signal that one is still present. Instead of a physical button press the conductor could confirm the alertness via Pursuits. Either by being told to look at a certain object and pursue it for the time being to indicate their capacity to act. Or different eye movements are analysed to indicate the awareness without a designated action. Similar analyses are used to advise the driver of a car to take a break when eye movements indicate decreasing attention.

Authentication The three single use cases categorised under 'Authentication' are strictly speaking three design aspects of authentication with Pursuits in the real world. As gaze authentication would be applied like a one time password by *being told what to look at*. After analysing the moving objects in the scene and then subsequently looking at this object/those objects a user can authenticate. This procedure can be technically implemented by utilising the user's smart phone cameras to record the environment (back facing camera) as well as tracking the user's eyes (front facing camera). The method is then used to enable the user to authenticate e.g. for online banking on their smart phone whilst being on the go and without risking giving their password away to bystanders and shoulder surfers. In other cases a single physical target is provided once, allowing to authenticate a group of people at once within a localised area (*providing a target - authenticate locally*). For example by throwing a ball in the air before a lecture and of course instructing all students to pursue it, a lecturer can include all present students in an online discussion board or connect them to a closed, local Wi-Fi network without having to declare a key.

Establishing Contact As eye contact is naturally used for making contact in several scenarios, use cases exploiting this behaviour and supporting it with technology were developed. Such as *informing the waiter at a restaurant*. In this case, the participants of the focus group identified that gaze was used to look for the waiter and follow them in order to catch the right moment to signal them. This process is supported using Pursuits. Rather than having the visually pursuing function as the set up to establish contact, with Pursuits it would be the establishment already. The user's eyes are monitored while looking at their waiter and after the movements are matched to a specific waiter, they are informed that guest at the respective table want service. Based on the same idea but with a more entertaining background is the idea behind the *dating event 'Ice skating'*, there one follows the person they are interested in and if that interest is shared both parties get a notification with contact information.

On big *construction sides* with many workers staying on top of things can be challenging. Pursuits can support supervising staff when establishing contact between single workers for example to inform them of changes or warn them. As opposed to collecting and searching for contact information for one person, Pursuits is used to identify the person of interest and set up a call or directly connect two radios with each other.

Entertainment Many of the concepts behind the ideas of categories above can also be applied to the entertainment sector. These should be see as a fun way to adapt Pursuits. Such as *insect laser* which is based on *determine target* (see *Hand-free Interaction/Selection*) but uses the selection in a very playful way. Annoying insects are eliminated after selecting them with Pursuits and are neutralised spraying them or, once available, use laser. The target set this way can also be used as input information for subsequent processes like in the scenario *muscle stimulation*. The information provided by the moving target is used to support muscle reaction, this way a tennis player for example can perfect their timing.

Games can also be designed to especially make use of Pursuits. For example in the adapted version of the *staring contest*. The user no longer has to win the fight of who can look at the other person the longest, but who can follow another person's movements the longest. Comparing the movement of one's finger to the users gaze makes cheating impossible and monitoring easier. Or adapting the wire loop game to gaze a game where the user has to *follow a trajectory without losing focus* otherwise they have to start over again. A different approach is need for the use case of *dancing arcade*, instead of having to follow a moving object, the user is asked to keep looking at a stationary object whilst moving, over the eye movements a system is now able to determine how the user is moving and give instructions for example the next dance move (also see section 2.1.4 Optokinetic Nystagmus and Vestibulo Ocular Reflex).

The previous ideas cover the game aspect of entertainment, but Pursuits can also be used as a tool of *community voting* to for example influence the plot of a film. Depending on which object the most viewer followed for instance during a scene of a crime film, the investigator will take a different lead to satisfy the users' curiosity.

Technical Applications One approach that has already been categorised with a more special concept, can also be used to determine the movement of a user (*have fixed point to determine the user's movement*). This can come in handy when the position of a user in a room is relevant but cannot be traced directly. Similar to this Pursuits can be used to calibrate a group of user's gaze toward a specific object and ensuring that everyone keeps at it by providing feedback (also used to authenticate a group with one object as explained in *Authentication*). Gaze data can also be used in VR to specify the level of detail of a scene depending on the object the user is looking at.

4.2 Design Space - Research Questions

Based on the scenarios of the brainstorming session some possible research questions are identified. These serve as ideas for future research and are not further investigated for now. The aspects can be grouped in three categories *system oriented*, *user oriented* and *target oriented* but are often depended on one another.

System Oriented The system in this context describes the entirety of components involved in the process of Pursuits, conceptual as well as technical. An interesting point for investigation can be the quality of results depending on the source of movements. We touched upon that in our case study (see section 4.3 Usability Study et seqq.) but did not solely focus on the technical aspects. Whether the target, the user or both are in motion can affect the performance but there might also be design aspects that can overcome these differences. Not only does the state of the single parties influence the outcome but also condition the requirements for the technical components and the data processing in the background (synchronising sources, translating coordinate systems etc.).

Another point to explore is how the system set up affects performance and user approval. As the technical devices can be specially designed (e.g. mobile eye trackers that need to be purchased) or are part of a already available system (e.g. front- and back-facing cameras of a smart phone, which might be an option to trace gaze in future). If the user does not themselves carry around the monitoring gear then the devices need to be set up on the scene for example with a stationary eye tracker at a restaurant table. An important aspect necessary to resolve in order for such scenarios to work is how to process the gaze data and relate it to information about the object. Movements might not be recorded from the same perspective therefore translations into a mutual coordinate system is necessary.

Target Oriented A couple of questions arise looking at different aspects of the targets. First of all the number of targets. Does the scene contain more than one moving object, so can false detections be made or is it merely a matter of detecting the moment the eye movement correlates to the stimulus' movement. Then the size of the target, the more area a stimulus provides for the user to look at, the more inconsistent their eye movement might become as the pursuit movement is interrupted by saccades. Then an approach considering shorter as well as longer time windows might be necessary to be able to grasp the direction of the movement.

This leads to the question how to determine the target area. In our study evaluations we use image processing on a manually coded black and white image afterwards and only return the centre of the round target. For interaction applications do not only require an interpretation of the data on the spot but also a more complex strategy to determine the position of the target. Certain objects might provide a natural point of interest the user naturally keeps looking at which can be used to calculate a single trajectory. If so how can targets be designed to draw the user's attention to one point. For other targets creating a set of trajectories based on single pixels and the centre of the object area might be an option. And how does size change affect the results, when for example an object first appears small in the distance and when closing in increases surface-area and level of detail.

We already addressed that target and user information could be gathered from different devices. But an interesting question is if different types of information can be combined and what these kinds of information are in regards to the target. For now we assumed we have two dimensional video material to observe movements as well as gaze position in 2D. But is it possible to map two dimensional gaze data with three dimensional object data obtained with depth sensors.

	A: Restaurant (user sitting)	**B: Restaurant (user moving)**	**C: ATM**
User	Stationary	Moving	Moving
Target	Moving	Moving	Stationary

Table 4.1: Usability study: state of target and user for each scenario

Or if QR-marker and NFC-tags can be used to support target detection and positioning in the room. And on the other hand is it possible to estimate a gaze positions in 3D, so mapping can be archived over three dimension for both parties.

A feature not only relevant to Pursuits in the real world but also on screen is, the way the target moves. One can distinguish between whether the motion motion periodically, repetitious or arbitrary. Or if it is predictable due to knowledge of the scene (e.g. street signs will always appear on the sides of the street and move alongside it) or because it keeps repeating itself. Is it somewhat predictable because the target is path bound (a waiter in a restaurant) or not at all predictable (a football player during the game). It might also be worth investigating if the performance is affected by how and when the target appears. A target entering the scene from the side of the user's field of view is potentially a better target than objects approaching frontally. The question that is especially relevant for real world scenarios is how users deal with periodically occlusion of the target, e.g. the animals at the zoo pass a tree. How well is the user at interpolating such gaps or how might Pursuits even be able to help.

User Oriented The user themselves also provide a couple of interesting points to look at, such as how the number of users influences the performance and how the gaze data of more users is organised. If the source of the object's data is different from the user's field of view, is this affecting the performance or does the user have to adapt in order for it to work. The more interesting aspects concerning the user however are user experience and usability oriented. Such as designing ways to introduce the method to potential user or making the use feel natural.

4.3 Usability Study

As our pre-study showed that performing test for Pursuits on real world scenarios with a video prototype might not return the same results, we designed a usability study as a wizard of Oz experiment in the real world. In order to get a first idea of how users would react to a Pursuit-based gaze interaction with real world stimuli instead of digital stimuli, we designed the study based on two use cases. Besides the user opinion we were also interested in how Pursuits would perform with regards to the movement state of target and user.

4.3.1 Usability-Study: Study Design

To investigate applicability and usability of using smooth pursuit as a mode of gaze interaction with the real world we designed a within subject repeated measure wizard of Oz experiment. Making use of the *inform waiter at the restaurant* scenario from the category *establishing contact* (see 4.1, Establishing Contact) developed during the brainstorming session. We developed a second scenario based on the ideas of the *authentication* category (see 4.1,Authentication). We chose and designed three scenarios presented to the user in such a way, that we can test the influence of movement states of target and user (see table 4.1) as well as the performance and usability. Scenario (A) restaurant (user sitting) adapts par to par the idea of the focus group. The user is presented with the situation in a restaurant where they try to signal the waiter that they want to order or get the bill. The user is hereby sitting as they visually follow the waiter as the target moving on a predefined path. Use case (B) restaurant (user moving) only differs in that the user is also moving e.g. on their way back from the buffet. The user's signalling is accomplished by simply pursuing the waiter with their gaze, performing the smooth pursuit movement. While during (A) and (B) the target is moving for (C) ATM the targets were stationary. We asked the user to approach a fabricated ATM but instead of using a PIN-code to authenticate, look at different targets when approaching the ATM.

For all three scenarios we introduced the sub-condition *head movement* with the variation *natural* and *still*. Allowing the user once to turn their head as desired in order to follow the target and a second time to keep their head as still as possible and only follow the movement with their eyes as long as the target can be well perceived only turning their head once the target had reached the outer zone of their field of view.

In addition to *head movement* for the two restaurant scenarios (A and B) we also varied the *feedback* given by the waiter between *visual feedback*, in which case the waiter would turn to the user and nod. And the second type, *audio feedback*, where the user heard an audio signal once the waiter registered the user's selection. Leading to four repetitions per restaurant scenario and two of the ATM scenario. Summing up to ten runs. The conditions were balanced using a latin square (see Appendix B).

4.3.2 Usability-Study: Procedure and Study Set-up

The study took place in two rooms within the university facilities where all obstacles were removed so that the user could walk safely as well as not be distracted by way finding. During the entire study the user was wearing the same mobile Pupil[8] eye tracker in their monocular form equipped with one eye camera and one front facing/world camera, that was used as during the pre-study. For scenario (A) Restaurant (user sitting) the user was sitting at a table observing the waiter in the room with the eye tracker being connected to a laptop via USB cable (see fig. 4.2 (a) and fig. 4.2 (c) without chairs for a reference of the walking area). For scenario (C) ATM, numbered chairs were arranged in the room to work as stationary targets (see fig. 4.2 (c)). The user walked in between the objects followed by a carrier of the laptop (see fig. 4.2 (b)). The latter was necessary to ensure the transmission of the camera feeds was not interrupted or terminated. As scenario (B) restaurant (user moving) presents a combination of the scenarios explained prior, the procedure was also combined. The waiter would again walk on a predefined path within the room being visually followed by the user, who walked on their own predefined path.

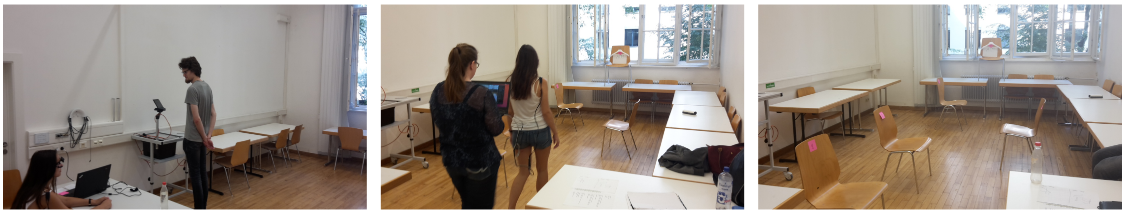
After every scenario the participant was asked to fill out the NASA TLX [14] questionnaire online[9] and a designated questionnaire (see Appendix C,D, E) also provided online, where they could judge the method on 5-Point Likert-scales and leave comments for the individual scenarios and sub-conditions. The session ended with a final questionnaire, also online (see Appendix F).

4.3.3 Usability-Study: Participants

We invited twelve participants to our one-hour long study, four female and 8 male (33%/66%). All of them were university students between the age of 19 and 35 of varying programmes. They had different levels of experience with eye tracking in general, e.g. previous studies that used eye tracking as a recording medium or dedicated eye tracking studies. However few had little experience with Pursuits (see fig. 4.3).

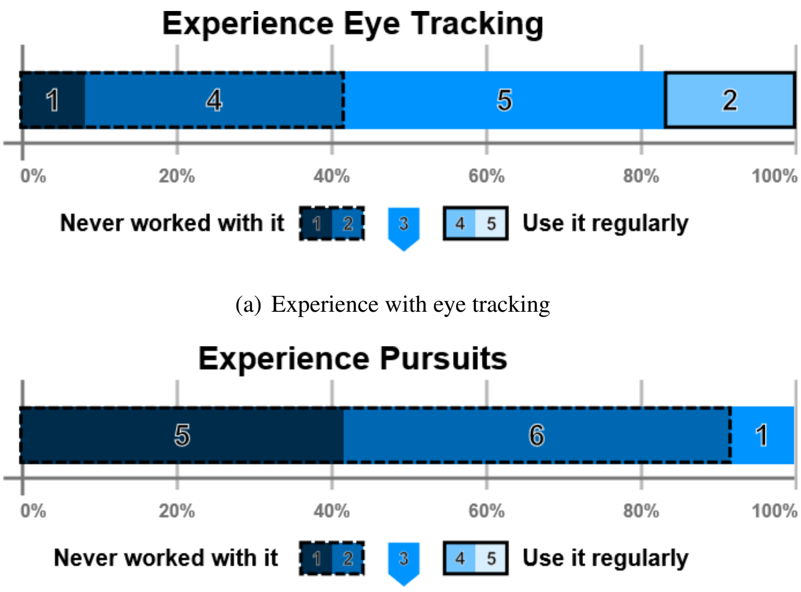
[8] pupil-labs.com

[9] www.keithv.com/software/nasatlx/



(a) Restaurant (user sitting) (b) ATM (c) ATM Set-up

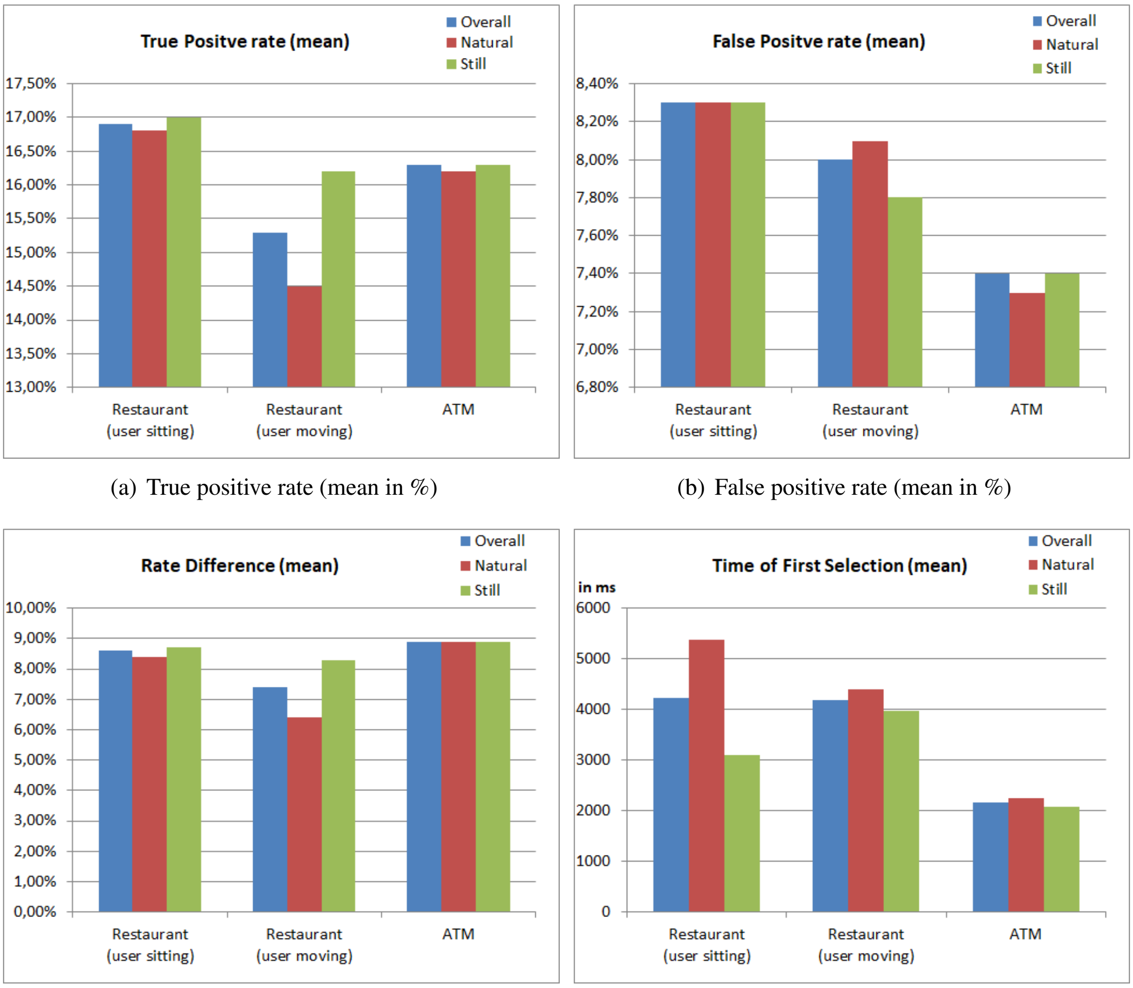
Figure 4.2: Usability Study: Study set-up



(a) Experience with eye tracking

(b) Experience with Pursuits

Figure 4.3: Level of experience. 60% of the participants were familiar with eye tracking technology, but only few experience with Pursuits.



(a) True positive rate (mean in %)

(b) False positive rate (mean in %)

(c) Rate difference (mean in percentage points displayed with %)

(d) Time of first selection (mean in ms)

Figure 4.4: True positive rate for the three scenarios.
Overall mean, natural head movement (mean), still head movement (mean)
Using window size = 500 ms. No significant differences can be reported for the use of Pursuits in different scenarios. Having move both, target and user, deems to be technically most challenging.

4.4 Evaluation & Results

In order to asses Pursuits in real world scenarios we evaluate the technical performance based on the same aspects we used to the pre-study applying the same process for the three different scenarios. Additionally we investigate usability and user experience by evaluating the NASA TLX [14] outcomes and the answers of the questionnaire. Starting with the technical analysis.

4.4.1 Usability-Study: Technical Evaluation & Results

To evaluate the technical performance of Pursuits in a real world based scenario, the *true positive rate*, *the false positive rate*, *the rate difference* and the *time of the first selection* (see section 3.3.1 Evaluation Aspects) are discussed for the three scenarios themselves as well as the influence of *head movement*. As the pre-study suggested using smaller window sizes we applied Pursuits with a window of 500ms.

True positive rate Viewing the true positive rates of the three scenarios (see fig. 4.4(a)) it is apparent that the performance of Pursuits decreases if both user and target are moving such as in the *Restaurant (user moving)* scenario (mean: 15.3% SD: 2.2%). Considering just this scenario the performance can be increased by keeping the head movement to a minimal, resulting in a plus of 1.7 percentage points. This pattern can be seen for all scenarios however not this distinct. Running a two-way repeated measure ANOVA these improvements turn out not to be significant. Likewise are the differences between scenarios and thereby the states of movement of target or user not significant in regards to the true positive rate. The classical approach with stationary user and a moving target results in slightly better rates (overall mean: 16.9% SD: 2.0%) than following stationary targets when walking (overall mean: 16.3% SD: 4.75%), as already stated both outdo the combination of moving user and target.

False positive rate Figure 4.4(b) shows that the false positive rates produced under the *restaurant (user sitting)* are equal for all *head movement* types, but are also the highest (overall mean: 8.3% SD: 0.7%). This is almost 1,0 percentage point higher than the condition with the least false positives (ATM overall mean: 7.4% SD: 0.9%). For *restaurant (user moving)* a controlled head movement also leads to better results for the false positive rates producing less false positives with no head movement (mean: 7.8% SD: 1.0%) than with head movement (mean: 8.1% SD: 1.0%). These minimal divinations again turn out not to be significant as well as the differences between scenarios.

Rate Difference Due to also producing high false positive rates, high true positive rates of well performing scenarios are compensated, resulting in a balances out rate difference. Only with *restaurant (user moving)* producing a perceptibly lower difference on the detailed diagram (see fig. 4.4(c). However running an ANOVA reveals that these differences are to small in order to have significant impact. Also differences between iterations with or without head movements yielded no statistical significant results for the rate differences. For the technically challenging situation of moving user and target again an adjusted head movement leads to better results (mean natural: 6.4 percentage points SD: 2.1 percentage points, mean still: 8.3 percentage points SD: 1.7 percentage points), but outcomes of the other scenarios expose no differences between *head movement* types.

Time of First Selection Other than for the evaluation aspects above when considering the time of the first selection the scenario with stationary targets seems to produce true positives the quickest (see fig. 4.4(d)). This low average however is due a lot of cases with no selection at all. For the ATM scenario 180 of 456 cases never resulted in a selection. For the otherwise technically weak *restaurant moving* only 70 of the 456 cases never ended in a selection and for *restaurant (user sitting)* only 32 cases terminated with no selection. This becomes apparent in the selection time when we correct the raw data to list a high entry for no selections and calculate the median instead of the mean (see fig. 4.5). This adapted data shows that for 50% of the cases a correct selection is made after 500 - 5000ms for the *ATM* scenario. Of the other 228 cases 180 never selected once, leaving only 48 cases in which the user would receive a response, the latest selection is made after 10.5 seconds. User who get a response receive it after an average of 2166.67ms. A first selection is made faster when the user actively moves their head, not only in the ATM scenario but in general. Even though true positive rates are lower when target and user are moving, the first selection is made in between 500ms and 2000ms in 50% of the cases and after 4182,98ms on average (if at all; maximum: 49.5 seconds). Only being minimally fast than the average selection of 4230,06ms for the *restaurant (user sitting)* situation, with a median of 3000ms (minimum: 500ms, maximum: 20 seconds). Two-way repeated measure ANOVA calculated no significance in these differences.

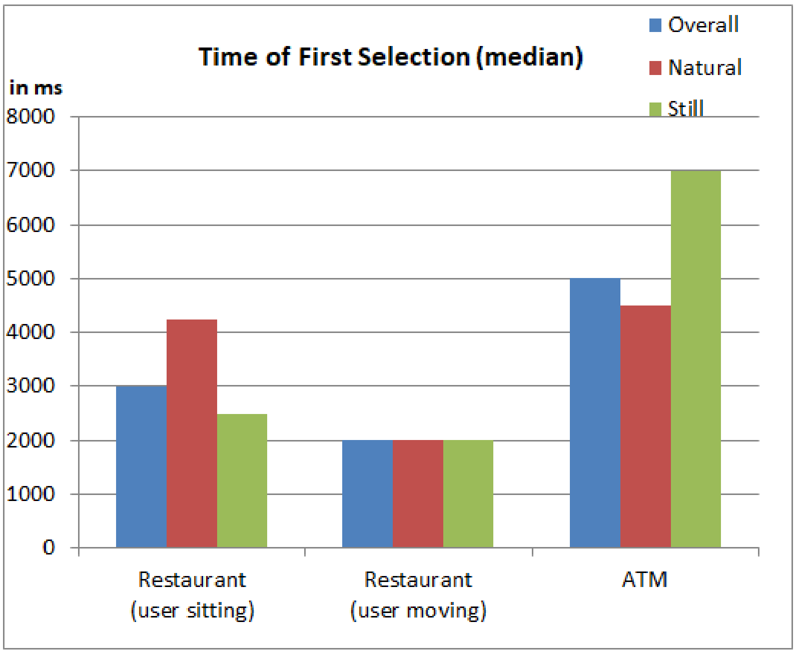
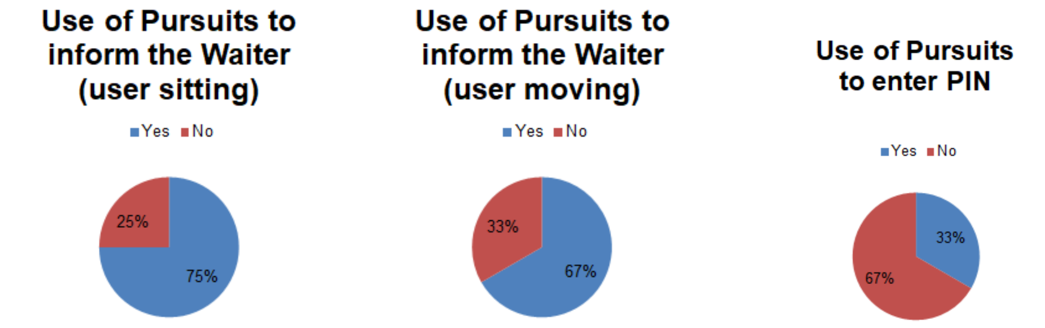


Figure 4.5: Time of first selection. Median in ms
A lot of non-selections alter the average of the selection time. Looking at the median reveals that selections made during the ATM scenario are overall slower than for the other scenarios.

4.4.2 Usability-Study: Usability and User Experience Evaluation & Results

In order to find out if and when users would want to interact using Pursuits-based gaze interaction, we presented a questionnaire (see Appendix C,D, E)) and asked them if it is likely that they use the approach they just tested. Only a quarter of the participants would not use it for the scenario where they were sitting and the waiter was moving around (see fig. 4.6). Equally high is the acceptance for the second restaurant scenario (still only one third that would not use it). For the ATM scenario however two third say they would not use Pursuits to enter a PIN this way. This can also be seen comparing the results of the Likert-scale rating. The participant could rate how much they liked the use of Pursuits on the previously performed scenario. For the scenario that most respondents would also use more almost 80% *liked* or *very much* liked using Pursuits (see fig. 4.7(a)). An equal picture is presented for the second restaurant scenario where the users would also walk around. In comparison to the first restaurant scenario only two people less rate the method with *liked* or *very much* liked (see fig. 4.7(b)). Even for the unpopular *ATM* scenario the minority disliked the approach utterly.

The reasons why people would use Pursuits in a restaurant setting are plenty. Most agree that it is a natural approach to contact the waiter as making eye contact is necessary anyway in order to catch the moment to signal the waiter. Two participants highlight the non-verbal aspect of this procedure as it can get quite loud in restaurants (stated by participants 5 and 6) and also enabling mute people to use it (stated by participant 5). A somewhat related aspect, that is pointed out, is that it still functions over greater distances and can quickly lead to success if well-engineered (stated by participant 7). The users who reject the system argue the approach to them it feels unnatural and verbally signalling the waiter is more comfortable (stated by participant 4 and 11). Furthermore it feels socially awkward or can be culturally unacceptable to follow someone visually for that long or turn around after somebody, which users could often be observed doing when they had to walk (stated by participant 12). One participant states that in order to gather the information necessary for the algorithm to work, the waiters of a restaurant have to be constantly

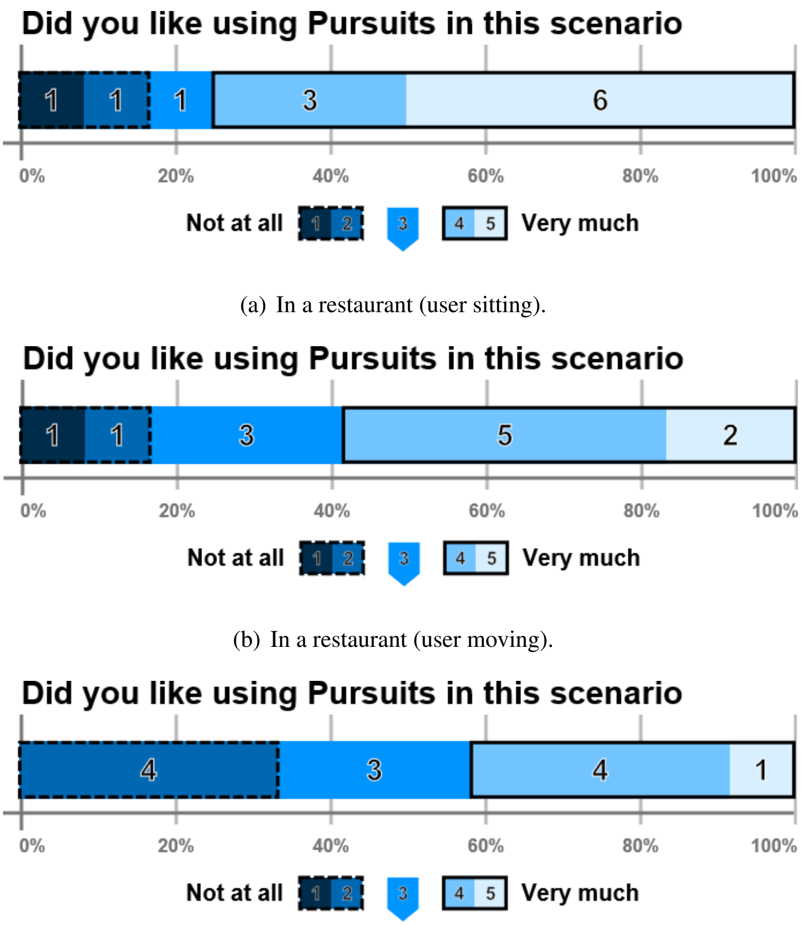


(a) In a restaurant (user sitting). (b) In a restaurant (user moving). (c) At an ATM

Figure 4.6: Would you use Pursuits in the following scenarios? - Most of the users would use Pursuits in cases where gaze interaction is a general means of interaction.

traced in addition to be looked at by people permanently. What in the eyes of the participant feels like a form of surveillance (stated by participant 3). Opposed to the predominantly user experience related comments on the use of Pursuits during the restaurant scenarios, the feedback for the *ATM* scenario is rather technical. User stated that they would not use it (for now) as they do not see a way to correct mistakes or how the approach can work if there is a queue at the ATM-machine and they are not walking towards it (stated by participant 2 and 10)). Many also have the impression that is it not secure and the head movements can give away the target and thereby the PIN (in that form stated by participant 7). On the other hand participants explain that they would use this approach in a one-time-password way, possibly in combination with augmented reality (AR) so they would not be fazed by 'shoulder surfing' the gaze password (stated by participant 5). Even more some have the feeling entering a PIN this way can be more secure and they would not have to remember a numerical PIN (stated by participant 4 and 5). The overall consensus for the usage of Pursuits to enter a PIN nevertheless is more negative as the users not only have doubts regarding the execution but also feel uncomfortable when using it and rate the approach impractical.

The most contradicting statement is that the use of Pursuits is unnatural and at the same time natural. On a Likert-scale from 1-5 we let the user rate how natural they found Pursuits in the respective scenario. Figure 4.8 shows that the scenario very much influences how Pursuits in perceived. In a setting like the restaurant, where according to the participants, eye contact is already naturally used, applying Pursuits also feels more natural. Over 50% of the participants rated the use on the natural side of the scale (see fig. 4.8(a)). It becomes unnatural when the user is also walking around (see fig. 4.8(b)) with only three participants rating the method as other unnatural. Although during the study it seemed that the participants found the scenario itself also unnatural. Signalling the waiter while they themselves are walking around seemed far fetched. The waiter would for example have no information regarding the table number nor did they see why they would signal the waiter when they are not at their table. Therefore we can only conclude that users did not like the overall scenario rather than the use of Pursuits while target and user are in motion. Same applies to the *ATM* scenario. None of the users felt that using Pursuits in this situation is natural (see fig. 4.8(c)). For us it is apparent that the use of Pursuits feels especially natural when gaze is already a means of interaction, such as when waiting for the right moment to signal a waiter. And is less natural when gaze is not usually a means of interaction. Like in our ATM scenario, when one would usually not use gaze to interact. The results shift for a distinction between the application of Pursuits for the different *head movements*, where overall the approach is perceived less natural with a controlled head movement (condition *still*) than the respective general rating.

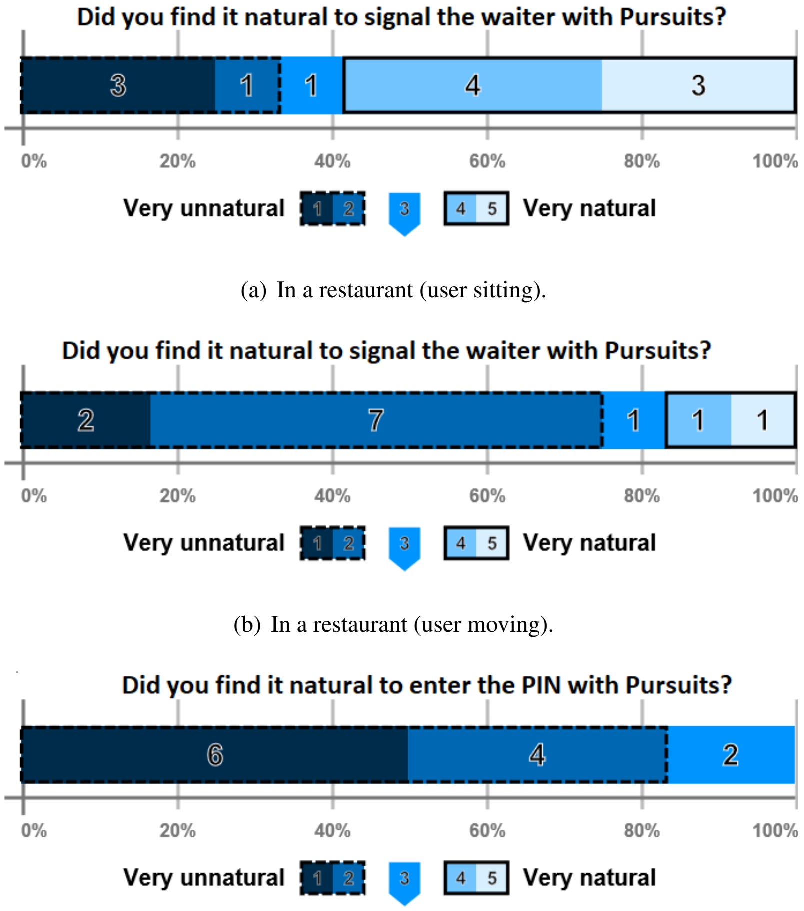


(a) In a restaurant (user sitting).

(b) In a restaurant (user moving).

(c) At an ATM

Figure 4.7: Did you like use of Pursuits natural in this scenario? - The users generally perceived Pursuits well and liked in most cases.



(a) In a restaurant (user sitting).

(b) In a restaurant (user moving).

(c) At an ATM

Figure 4.8: Did you find the use of Pursuits natural in this scenario? - Most of the users found the use Pursuits natural in cases where gaze interaction is a general means of interaction

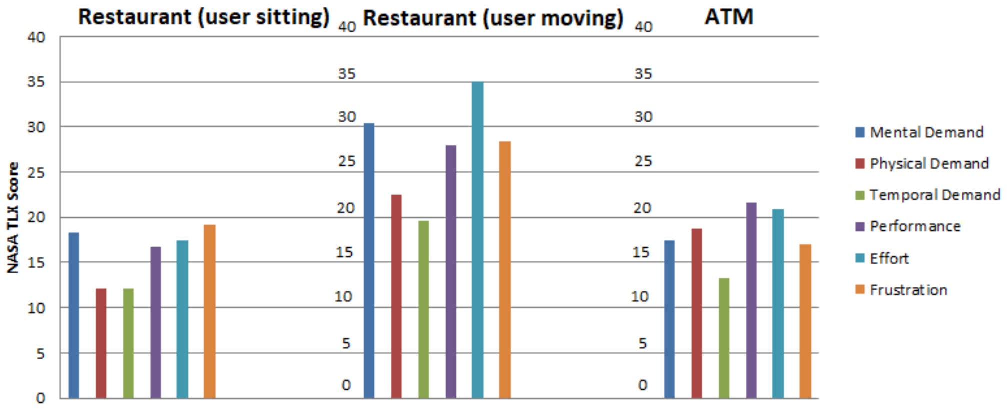
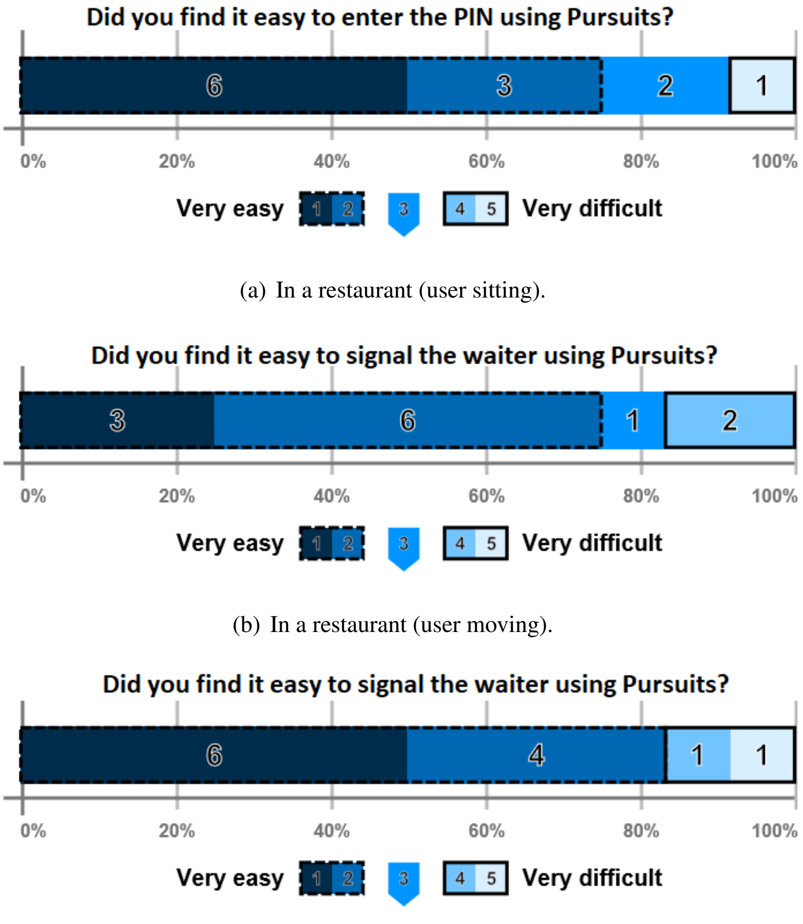


Figure 4.9: Results of raw NASA TLX. Restaurant (user sitting)/ (user moving), ATM. Restaurant (user moving) scenario is an overall more complex task (significant), however no significant differences between sub-scales could be verified.

To make the waiter scenario less distanced we gave different kinds of feedback during the study when a selection was supposed to be made. For two runs (once for each *head movement* sub-condition) per scenario the user would get personal feedback from the waiter (nodding) for the other two runs audio feedback with no reaction by the waiter was provided. In general the participants liked having feedback at all (only one person would prefer no feedback during the walking situation). A majority enjoyed the personal feedback more as it is ‘more natural’, ‘more personal’ and less ‘dehumanising’ to the waiter (stated by participant 4,6 and 3). Besides sating environment-related aspects like audio would be annoying to other guests or can be over-heard (stated by participant 9). A nod would be naturally logically be seen, as the user is already looking at the waiter in order to perform Pursuits.

For assessing the workload during the scenarios we had the participants fill out the NASA TLX [14] questionnaire online[10]. We learnt that Pursuits does not perform as well for the scenario with moving user and target, technically speaking and also the users seem to struggle with that one. Users rate the scenario higher on all sub-scales, especially values for *mental demand*, *physical demand* and *effort* almost double comparing to the other restaurant scenario or the *ATM* situation. Two-way repeated measure ANOVA also reveals significant differences between scenarios for *mental demand* ($p = .04$) and *physical demand* ($p = .037$). But Bonferroni-corrected post-hoc tests could not confirm significant deviations between the scenarios. Comparing the overall work load the NASA TLX measures, the same behaviour can be observed. With a overall score mean of 30.633 the walking condition of the restaurant scenario gets almost double the scoring of its stationary counterpart (overall score mean: 18.361) which ANOVA with a post-hoc Bonferroni yields as significant (ANOVA: $p = .006$, Bonferroni: $p = .021$). The scoring for the ATM scenario is not significantly different compared to both other scenarios. This observation does not match the users perception though. For both restaurant scenarios almost 80% rate the use *easy* and *very easy* only varying in the exact amount between those two categories. Just the *ATM* scenario is rated better (over 80% for *easy* and *very easy*) which on the other hand is not rated neutrally but only *difficult* and *very difficult* by the remaining somewhat 20%.

[10] www.keithv.com/software/nasatlx/



(a) In a restaurant (user sitting).

(b) In a restaurant (user moving).

(c) At an ATM

Figure 4.10: Did you find the use of Pursuits easy in this scenario? - Overall the participant perceived Pursuits interaction with the real world as easy.

5 Discussion

Comparing Pursuits for different window sizes. With our pre-study we analysed the performance of Pursuits on a real world stimulus presented on three different visual environment. We found out that for movements executed in the real world short time windows lead to better outcomes of Pursuits regardless of the environment. With high *true positive rates* of 20.4% correlating real world gaze data to the object's movements. 17.6% when correlating it with gaze data produced following the stimulus on an unaltered video. And 17.4% for gaze data collected while following the abstract stimulus on-screen. Under this window size Pursuits also produced low *false positive rates* in relation to the respective high *true positive rates* resulting in high *rate differences*. A 11.5 percentage point rate difference is archived with *real world* gaze data. For *original video* data the *true positive rate* is 8.8 percentage points better and 9.2 percentage points occur between rates when relating *abstract video* data. The use of the 500ms window also enable selections to be made quicker resulting in average selection times of 2764.34ms, 2393.06ms and 2582.35ms (*real world*, *original video*, *abstract video*).

Comparing Pursuits for different environments. Because of this outcome we compared the performance of Pursuits on the three display different environments *real world*, *orignal video* and *abstract video* using a 500ms window. Statistical analyses showed that the differences produced under different environment conditions for all evaluation aspects are not significant. But ratings for the conduct of the test with a real world stimulus in the real world vary and are interestingly higher compared to the two video environments (e.g. 20.4% true positive rate compared to 17.4% and 17.6%). On account of that and user comments we received despite not conducting any surveys we suggest testing real world Pursuits scenarios not with digital prototypes but in the correct setting, in order to get the most accurate results.

Analysing the technical performance of Pursuits for use cases. Minding our own suggestion we designed a usability study for three real world use cases as a wizard of Oz experiment within the true setting of a three dimensional room instead of on-screen. The scenarios were chosen so that we could cover all motion state combination of user and target: the 'classical approach' with a moving target and a stationary user (*restaurant (user sitting)*), the inverse situation with a moving user and stationary targets (*ATM*) and the mix of both with moving user as well as target (*restaurant (user moving)*). The Pursuits performance analysis showed that having move both parties is technically the most challenging approach producing an however insignificantly lower *true positive rate* of 15.3%. Additionally a low *rate differences* of 7.4 percentage points compared to the 'classical approach' (16.9%, 8.6 percentage points) and the *ATM* scene (16.3%, 8.9 percentage points). But we also found that for the ATM scenario a bigger amount of cases resulted in no selection at all than for all other scenarios. 39.47% of non-selected cases can be detected for the *ATM* scenarios, which is 63.82% of the over all cases that ended with no selection.

Usability study design flaw. This circumstance is possibly due to (the design of) the scenarios. With the way Pursuits is evaluated, based on the gaze data and object data obtained from a video, recorded by the front-facing camera of the mobile eye tracker, results heavily rely on the camera angle. The recorded material of the scenario would often not correspond with the users field of view as this is much bigger than the camera angle. If the user was now to look down at a relatively close target without moving their head, the camera most likely did not pick up the area they were just looking at. Adapting the head position to match the current field of view seemed especially unnatural for the *head movement* condition *still*, as the user would not approach an ATM machine with a lowered head.

This can also be observed when looking at the data where the non-selection rate is slightly higher for the runs with no head movement (53.33% of the non-selection of ATM are allotted to the *still* condition). This issue did not occur for the restaurant scenarios as often, since most of the target's movement happened within the centre of the field of view, which is well covered by the camera image.

Assessing the usability of Pursuits in the real world. For this second study we aimed to gather information about the usability and the user experience of using Pursuits as an interaction method with the real world. Therefore we carried out a user survey including a work load assessment using the NASA TLX. Evaluating our self-designed questionnaire we learnt that the users are open to the idea of applying this kind of gaze interaction to real world scenarios. On the condition that gaze is already a natural or intuitive/trained means of interaction. So was the acceptance for Pursuits much higher in the use case where gaze was used to interact anyway, such as when trying to catch the right moment to signal the waiter. Than when gaze is usually not part of an active interaction, for example when entering a PIN at the ATM. This circumstance also shows in the received ratings for naturalness which were mostly positive for the first restaurant scenario but mainly and exclusively negative for the walking restaurant scenario and the ATM scenario. Although from comments during the study we can presume that users did not dislike using Pursuits while walking (on moving or stationary targets) but rather disliked the situations themselves. Mostly it was unclear why a restaurant guest would need to contact a waiter non-verbally while walking. Either people of the group remain at the table can signal the waiter or the walking person would run into a waiter randomly on their way to e.g. the restrooms and be able to communicate that guests at their table need service. But why someone would want to signal the waiter from a distance when being away from their table seemed incomprehensible. For the ATM scenario comments where of rather technical nature: pointing out that a means for correcting wrong input is missing as well as communicating security concerns.

6 Future Work

A list of potential research questions is already compiled in section 4.2 Design Space - Research Questions. Containing aspects regarding the use of different information in order to calculate Pursuits. Thereby we suggest investigating whether the three dimensional information collected by depth sensors can be used to correlate with usually two dimensional gaze data. In addition to that already mentioned feature we propose to explore the option of measuring the depth of user's gaze [30]. Potentially important attributes can be the convergence of the eye which can be estimated based on the data of e.g binocular eye trackers or EGO measurements. This way the correlation on all three dimension could be compared. In the section referenced above we also suggest enhancing target tracking by other means such as NFC-tags or visual marks. The data synchronisation, feasibility and effectiveness has to be examined.

In this work we only superficially tested the effect the movement state of target and user in the scope of an usability study. Our study showed that using targets that are only in motion from the user's perspective as they are moving works in generally and offers the possibility for several applications (see 4.1 Brainstorming Session). But we also highlighted that the position of these targets can influence the quality of Pursuits if a video-based approach is used. A dedicated study determining the potential stationary targets have and an ideal design of these scenarios should be conducted before application. Also investigating weak performing trajectories that are possible in three dimensional space in more detail is recommend. We touched upon that subject with our pre-study, could, however, not single out trajectories yet, as our emphasis was to discover differences in the displaying environment. Early analyses show that very frontally nearing trajectories for example lead to few eye movement and worse results. A closer inspection of that circumstances can help eliminate use cases with smaller chances of success.

Closely related is the question whether objects on the same path vary enough to produce different trajectories. Imagining the use case at a factory or a sushi restaurant (see 4.1 Brainstorming Session, Information) where all objects move in line, depending on the design of the path objects share a trajectory for longer periods of time, in order for Pursuits to still work other aspects might need to be considered. An even more extreme case is presented with our tested ATM scenario. All targets were positioned left and right from the user essentially presenting the same trajectories, in our evaluation we only coded one target at the time depending on the user's point of interest. In a potential automated object recognition process however all targets would be present at the same time. Whether this is a defeat of Pursuits or if other properties can be considered needs to be investigated.

As already mentioned a long time goal would be to apply Pursuits in real time. Therefore automated object recognition needs to be implemented that can identify and locate potential targets. A first step would be to realise automated post-hoc techniques in order to automatise a similar evaluation methodology to the one used above. Only the last step would be compiling a system that would track gaze at the same time it identifies targets and correlating them with no delay.

Another aspect that needs attention in future, is how to provide feedback to the user for a successful or also failed selection. In our usability study we touched upon that by providing audio or personal feedback of the waiter and learnt that this is an important issue for users. Whereas audio feedback was less well-received than personal feedback, which corresponds to others' findings [47]. Alternatively other visual or haptic feedback modalities are conceivable [23].

A non-gaze-based but also correlation-based approach is path mimicry [5]. In order to interact with devices by using low-cost technology gestures matching a displayed motion are executed and correlated [6]. Translating this approach to arbitrary real world objects can offer further, also gaze-combined, interaction techniques.

7 Conclusion

As exploiting the smooth pursuits eye movement with Pursuits is well established for on-screen application we suggest translating it to the real world. Enabling users to interaction with any every day item they desire. Provided that this item is moving in relation to them, either because itself is in motion or because the user is moving past it.

We conducted a first data driven pre-study in order to detect differences in the performance of Pursuits with real world stimuli on different display environments. Therefore we showed stimulus movements in the real world to users and had them watch these movements once in the real world, once as the unaltered video and once with an abstracted video on-screen. In all three cases their gaze was recorded using eye trackers so that Pursuits could be executed post-hoc on the gaze and object data collected during the study. Results showed that Pursuits generally works with real world stimuli and outcomes are better for smaller window sizes when applying Pursuits to real world scenarios independent of the used environment. Even though the evaluation did not result in significant differences, due to deviations in rates and user feedback, we conclude that using video prototypes for real world Pursuits scenarios is insufficient and varying results have to be expected.

In a further study we tested Pursuits in real world use cases. We therefore designed three scenarios based on use cases developed during a brainstorming session. The cases additionally covered all three combinations of user and target motion states: a sitting user and moving target, stationary targets and a walking user and the mixture. It turned out that the latter is technically the most challenging, producing weaker results than the other two. The users were generally very open for the idea of interacting with real world objects using Pursuit, under the condition that the application of gaze interaction is natural in the scenario. We suggest designing application for scenarios where gaze and especially the smooth pursuit movement is already used to interact even without technical support.

A Pre-Study: Latin Square

Linear Baseline	L-zAxis	C-Tilted	L-DiagonalPass	C-yAxis	L-Height	C-xAxis	Circular Baseline
L-zAxis	L-DiagonalPass	Linear Baseline	L-Height	C-Tilted	Circular Baseline	C-yAxis	C-xAxis
L-DiagonalPass	L-Height	L-zAxis	Circular Baseline	Linear Baseline	C-xAxis	C-Tilted	C-yAxis
L-Height	Circular Baseline	L-DiagonalPass	C-xAxis	L-zAxis	C-yAxis	Linear Baseline	C-Tilted
Circular Baseline	C-xAxis	L-Height	C-yAxis	L-DiagonalPass	C-Tilted	L-zAxis	Linear Baseline
C-xAxis	C-yAxis	Circular Baseline	C-Tilted	L-Height	Linear Baseline	L-DiagonalPass	L-zAxis
C-yAxis	C-Tilted	C-xAxis	Linear Baseline	Circular Baseline	L-zAxis	L-Height	L-DiagonalPass
C-Tilted	Linear Baseline	C-yAxis	L-zAxis	C-xAxis	L-DiagonalPass	Circular Baseline	L-Height

B Usability Study: Latin Square

WS				WM				ATM	
back		audio		back		audio			
nat	still	nat	still	nat	still	nat	still	nat	still
WS				WM				ATM	
back		audio		back		audio			
still	nat	still	nat	still	nat	still	nat	still	nat
WM				ATM		WS			
back		audio				back		audio	
nat	still	nat	still	nat	still	nat	still	nat	still
WM				ATM		WS			
back		audio				back		audio	
still	nat	still	nat	still	nat	still	nat	still	nat
ATM		WS				WM			
		back		audio		back		audio	
nat	still	nat	still	nat	still	nat	still	nat	still
ATM		WS				WM			
		back		audio		back		audio	
still	nat	still	nat	still	nat	still	nat	still	nat
ATM		WM				WS			
		back		audio		back		audio	
nat	still	nat	still	nat	still	nat	still	nat	still
ATM		WM				WS			
		back		audio		back		audio	
still	nat	still	nat	still	nat	still	nat	still	nat
WS				ATM		WM			
back		audio				back		audio	
nat	still	nat	still	nat	still	nat	still	nat	still
WS				ATM		WM			
back		audio				back		audio	
still	nat	still	nat	still	nat	still	nat	still	nat
WM				WS				ATM	
back		audio		back		audio			
nat	still	nat	still	nat	still	nat	still	nat	still
WM				WS				ATM	
back		audio		back		audio			
still	nat	still	nat	still	nat	still	nat	still	nat

WS				WM				ATM	
audio		back		audio		back			
nat	still	nat	still	nat	still	nat	still	nat	still
WS				WM				ATM	
audio		back		audio		back			
still	nat	still	nat	still	nat	still	nat	still	nat
WM				ATM		WS			
audio		back				audio		back	
nat	still	nat	still	nat	still	nat	still	nat	still
WM				ATM		WS			
audio		back				audio		back	
still	nat	still	nat	still	nat	still	nat	still	nat
ATM		WS				WM			
		audio		back		audio		back	
nat	still	nat	still	nat	still	nat	still	nat	still
ATM		WS				WM			
		audio		back		audio		back	
still	nat	still	nat	still	nat	still	nat	still	nat
ATM		WM				WS			
		audio		back		audio		back	
nat	still	nat	still	nat	still	nat	still	nat	still
ATM		WM				WS			
		audio		back		audio		back	
still	nat	still	nat	still	nat	still	nat	still	nat
WS				ATM		WM			
audio		back				audio		back	
nat	still	nat	still	nat	still	nat	still	nat	still
WS				ATM		WM			
audio		back				audio		back	
still	nat	still	nat	still	nat	still	nat	still	nat
WM				WS				ATM	
audio		back		audio		back			
nat	still	nat	still	nat	still	nat	still	nat	still
WM				WS				ATM	
audio		back		audio		back			
still	nat	still	nat	still	nat	still	nat	still	nat

C Usability Study: Questionnaire (Restaurant (user sitting))

Kannst du dir vorstellen mit Pursuits dem Kellner in einem Restaurant zu signalisieren, dass du z.B. bezahlen möchtest?
o Ja | o Nein | Begründe bitte deine oben gegebene Antwort kurz.

Wie gut hat dir das Verfahren gefallen?
Gar nicht 1 2 3 4 5 Sehr gut

ÄLLGEMEIN: Wie einfach war es, das Ziel in diesem Szenario zu selektieren?
Sehr einfach 1 2 3 4 5 Sehr schwer

ALLGEMEIN: Wie natürlich fandst du es in diesem Szenario, das Ziel so zu selektieren?
Sehr unnatürlich 1 2 3 4 5 Sehr natürlich

MIT KOPFBEWEGUNG: Wie einfach war es, das Ziel zu selektieren?
Sehr einfach 1 2 3 4 5 Sehr schwer

MIT KOPFBEWEGUNG: Wie natürlich fandst du es, das Ziel so zu selektieren?
Sehr unnatürlich 1 2 3 4 5 Sehr natürlich

OHNE KOPFBEWEGUNG: Wie einfach war es, das Ziel zu selektieren?
Sehr einfach 1 2 3 4 5 Sehr schwer

OHNE KOPFBEWEGUNG: Wie natürlich fandst du es, das Ziel so zu selektieren?
Sehr unnatürlich 1 2 3 4 5 Sehr natürlich

Du hast auf zwei verschiedene Weisen Feedback bekommen. Hat dir eine besser gefallen?
o Ja | o Nein

Kannst du dir Audio-Feedback oder Kellner-Signal besser in einem echten Restaurant vorstellen?
o Audio | o Kellner | o Beides gleich gut | o Beides gar nicht

Begründe bitte deine Antworten zu beiden Feedback Fragen kurz.

Sonstige Anmerkungen zu diesem Szenario:

D Usability Study: Questionnaire (Restaurant (user moving))

Kannst du dir vorstellen mit Pursuits dem Kellner in einem Restaurant zu signalisieren, dass du z.B. bezahlen möchtest?
o Ja | o Nein | Begründe bitte deine oben gegebene Antwort kurz.

Wie gut hat dir das Verfahren gefallen?
Gar nicht 1 2 3 4 5 Sehr gut

Hat dir das Herumlaufen und ein sich bewegenedes Ziel verfolgen Probleme bereitet?
Gar nicht 1 2 3 4 5 Sehr stark

ALLGEMEIN: Wie einfach war es, das Ziel in diesem Szenario zu selektieren?
Sehr einfach 1 2 3 4 5 Sehr schwer

ALLGEMEIN: Wie natürlich fandst du es in diesem Szenario, das Ziel so zu selektieren?
Sehr unnatürlich 1 2 3 4 5 Sehr natürlich

MIT KOPFBEWEGUNG: Wie einfach war es, das Ziel zu selektieren?
Sehr einfach 1 2 3 4 5 Sehr schwer

MIT KOPFBEWEGUNG: Wie natürlich fandst du es, das Ziel so zu selektieren?
Sehr unnatürlich 1 2 3 4 5 Sehr natürlich

OHNE KOPFBEWEGUNG: Wie einfach war es, das Ziel zu selektieren?
Sehr einfach 1 2 3 4 5 Sehr schwer

OHNE KOPFBEWEGUNG: Wie natürlich fandst du es, das Ziel so zu selektieren?
Sehr unnatürlich 1 2 3 4 5 Sehr natürlich

Du hast auf zwei verschiedene Weisen Feedback bekommen. Hat dir eine besser gefallen?
o Ja | o Nein

Kannst du dir Audio-Feedback oder Kellner-Signal besser in einem echten Restaurant vorstellen?
o Audio | o Kellner | o Beides gleich gut | o Beides gar nicht

Begründe bitte deine Antworten zu beiden Feedback Fragen kurz.

Sonstige Anmerkungen zu diesem Szenario:

E Usability Study: Questionnaire (ATM)

Kannst du dir vorstellen dich so am Geldautomaten anzumelden, statt den PIN einzugeben?
o Ja | o Nein | Begründe bitte deine oben gegebene Antwort kurz.

Wie gut hat dir das Verfahren gefallen?
Gar nicht 1 2 3 4 5 Sehr gut

Hat dir das Verfolgen von Zielen während des Laufens Probleme bereitet?
Gar nicht 1 2 3 4 5 Sehr stark

ALLGEMEIN: Wie einfach war es, das Ziel in diesem Szenario zu selektieren?
Sehr einfach 1 2 3 4 5 Sehr schwer

ALLGEMEIN: Wie natürlich fandst du es in diesem Szenario, das Ziel so zu selektieren?
Sehr unnatürlich 1 2 3 4 5 Sehr natürlich

MIT KOPFBEWEGUNG: Wie einfach war es, das Ziel zu selektieren?
Sehr einfach 1 2 3 4 5 Sehr schwer

MIT KOPFBEWEGUNG: Wie natürlich fandst du es, das Ziel so zu selektieren?
Sehr unnatürlich 1 2 3 4 5 Sehr natürlich

OHNE KOPFBEWEGUNG: Wie einfach war es, das Ziel zu selektieren?
Sehr einfach 1 2 3 4 5 Sehr schwer

OHNE KOPFBEWEGUNG: Wie natürlich fandst du es, das Ziel so zu selektieren?
Sehr unnatürlich 1 2 3 4 5 Sehr natürlich

Siehst du bei diesem Verfahren Vor- oder Nachteile gegenüber der Eingabe von PINs? Beschreibe diese bitter kurz:

Sonstige Anmerkungen zu diesem Szenario:

F Usability Study: Final Questionnaire

Alter:

Geschlecht:

Beruf/Studienrichtung:

Wie erfahren, würdest du sagen, bist du mit Eye-Tracking im Allgemeinen?
Noch nie damit gearbeitet 1 2 3 4 5 Verwende ich regelmäßig

Wie erfahren bis du mit Pursuits?
Noch nie damit gearbeitet 1 2 3 4 5 Verwende ich regelmäßig

Hast du noch Anmerkungen zu Fragen, deinen Antworten, der Studie, dem Verfahren . . .

Inhalt der beigelegten CD

\Thesis

LaTeX-File, BibTex-File, PDF, Figures

\Evaluation

\Pre-Study

.xlsx - Files used for the Evaluation of the pre-study
+ coded csv. - source files for evaluation

\Usability Study

.xlsx - Files used for the Evaluation of the usability study
+ coded csv. - source files for evaluation

\Software\Evaluation

\Merging Tools

Tools that merge gaze log and position log

\Single Participnat

Evaluation tool, version to evaluate only single Participants

\Pre-Study

Evaluation tool with hard-coded condition number to 8

\Usability Study

Evaluation tool with condition number variable included set to 6

\Software\Video player

Video player used in the pre-study

\Presentation

Slides used during the presentation

Literatur

[1] Anagnostopoulos, V. A., and Kiefer, P. Towards gaze-based interaction with urban outdoor spaces. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct* (2016), ACM, pp. 1706–1715.

[2] Bengoechea, J. J., Villanueva, A., and Cabeza, R. Hybrid eye detection algorithm for outdoor environments. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (2012), ACM, pp. 685–688.

[3] Bolt, R. A. Gaze-orchestrated dynamic windows. In *ACM SIGGRAPH Computer Graphics* (1981), vol. 15, ACM, pp. 109–119.

[4] Brignull, H., and Rogers, Y. Enticing people to interact with large public displays in public spaces. In *Proceedings of INTERACT* (2003), vol. 3, pp. 17–24.

[5] Carter, M., Velloso, E., Downs, J., Sellen, A., O'Hara, K., and Vetere, F. Pathsync: Multi-user gestural interaction with touchless rhythmic path mimicry. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2016), CHI '16, ACM, pp. 3415–3427.

[6] Clarke, C., Bellino, A., Esteves, A., Velloso, E., and Gellersen, H. Tracematch: A computer vision technique for user input by tracing of animated controls. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (New York, NY, USA, 2016), UbiComp '16, ACM, pp. 298–303.

[7] Cymek, D. H., Venjakob, A. C., Ruff, S., Lutz, O. H.-M., Hofmann, S., and Roetting, M. Entering pin codes by smooth pursuit eye movements. *Journal of Eye Movement Research 7*, 4 (2014).

[8] De Luca, A., Weiss, R., and Drewes, H. Evaluation of eye-gaze interaction methods for security enhanced pin-entry. In *Proceedings of the 19th australasian conference on computer-human interaction: Entertaining user interfaces* (2007), ACM, pp. 199–202.

[9] Drewes, H., and Schmidt, A. Interacting with the computer using gaze gestures. *Human-Computer Interaction–INTERACT 2007* (2007), 475–488.

[10] Esteves, A., Velloso, E., Bulling, A., and Gellersen, H. Orbits: Gaze interaction for smart watches using smooth pursuit eye movements. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (2015), ACM, pp. 457–466.

[11] Gowases, T., Bednarik, R., and Tukiainen, M. Gaze vs. mouse in games: The effects on user experience.

[12] Hansen, D. W., Skovsgaard, H. H., Hansen, J. P., and Møllenbach, E. Noise tolerant selection by gaze-controlled pan and zoom in 3d. In *Proceedings of the 2008 symposium on Eye tracking research & applications* (2008), ACM, pp. 205–212.

[13] Hansen, J. P., Tørning, K., Johansen, A. S., Itoh, K., and Aoki, H. Gaze typing compared with input by head and hand. In *Proceedings of the 2004 symposium on Eye tracking research & applications* (2004), ACM, pp. 131–138.

[14] Hart, S. G., and Staveland, L. E. Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Advances in psychology 52* (1988), 139–183.

[15] HEIKKILÄ, H., AND RÄIHÄ, K.-J. Speed and accuracy of gaze gestures. *Journal of Eye Movement Research 3*, 2 (2009).

[16] ISOKOSKI, P., AND MARTIN, B. Eye tracker input in first person shooter games. In *Proceedings of the 2nd Conference on Communication by Gaze Interaction: Communication by Gaze Interaction-COGAIN 2006: Gazing into the Future* (2006), pp. 78–81.

[17] ISTANCE, H., BATES, R., HYRSKYKARI, A., AND VICKERS, S. Snap clutch, a moded approach to solving the midas touch problem. In *Proceedings of the 2008 symposium on Eye tracking research & applications* (2008), ACM, pp. 221–228.

[18] ISTANCE, H., HYRSKYKARI, A., IMMONEN, L., MANSIKKAMAA, S., AND VICKERS, S. Designing gaze gestures for gaming: an investigation of performance. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* (2010), ACM, pp. 323–330.

[19] ISTANCE, H., HYRSKYKARI, A., VICKERS, S., AND CHAVES, T. For your eyes only: Controlling 3d online games by eye-gaze. In *IFIP Conference on Human-Computer Interaction* (2009), Springer, pp. 314–327.

[20] JACOB, R. J. What you look at is what you get: eye movement-based interaction techniques. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (1990), ACM, pp. 11–18.

[21] JACOB, R. J. Eye movement-based human-computer interaction techniques: Toward non-command interfaces. *Advances in human-computer interaction 4* (1993), 151–190.

[22] JALALINIYA, S., AND MARDANBEGI, D. Eyegrip: Detecting targets in a series of uni-directional moving objects using optokinetic nystagmus eye movements. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2016), CHI '16, ACM, pp. 5801–5811.

[23] KANGAS, J., ŠPAKOV, O., ISOKOSKI, P., AKKIL, D., RANTALA, J., AND RAISAMO, R. Feedback for smooth pursuit gaze tracking based control. In *Proceedings of the 7th Augmented Human International Conference 2016* (New York, NY, USA, 2016), AH '16, ACM, pp. 6:1–6:8.

[24] KHAMIS, M., ALT, F., AND BULLING, A. A field study on spontaneous gaze-based interaction with a public display using pursuits. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers* (2015), ACM, pp. 863–872.

[25] KHAMIS, M., HOESL, A., KLIMCZAK, A., REISS, M., ALT, F., AND BULLING, A. Eyescout: Active eye tracking for position and movement independent gaze interaction with large public displays. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology* (2017), UIST '17.

[26] KHAMIS, M., SALTUK, O., HANG, A., STOLZ, K., BULLING, A., AND ALT, F. Textpursuits: Using text for pursuits-based interaction and calibration on public displays. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (2016), ACM, pp. 274–285.

[27] KHAMIS, M., TROTTER, L., TESSMANN, M., DANNHART, C., BULLING, A., AND ALT, F. Eyevote in the wild: do users bother correcting system errors on public displays? In *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia* (2016), ACM, pp. 57–62.

[28] LUTZ, O. H.-M., VENJAKOB, A. C., AND RUFF, S. Smoovs: Towards calibration-free text entry by gaze using smooth pursuit movements. *Journal of Eye Movement Research 8*, 1 (2015).

[29] MAJARANTA, P., AND BULLING, A. Eye tracking and eye-based human–computer interaction. In *Advances in physiological computing*. Springer, 2014, pp. 39–65.

[30] MANSOURYAR, M., STEIL, J., SUGANO, Y., AND BULLING, A. 3d gaze estimation from 2d pupil positions on monocular head-mounted eye trackers. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications* (2016), ACM, pp. 197–200.

[31] MOLLENBACH, E., HANSEN, J. P., LILLHOLM, M., AND GALE, A. G. Single stroke gaze gestures. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems* (2009), ACM, pp. 4555–4560.

[32] PFEUFFER, K., ALEXANDER, J., CHONG, M. K., AND GELLERSEN, H. Gaze-touch: combining gaze with multi-touch for interaction on the same surface. In *Proceedings of the 27th annual ACM symposium on User interface software and technology* (2014), ACM, pp. 509–518.

[33] PFEUFFER, K., VIDAL, M., TURNER, J., BULLING, A., AND GELLERSEN, H. Pursuit calibration: Making gaze calibration less tedious and more flexible. In *Proceedings of the 26th annual ACM symposium on User interface software and technology* (2013), ACM, pp. 261–270.

[34] PFEUFFER, K., ZHANG, Y., AND GELLERSEN, H. A collaborative gaze aware information display. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers* (2015), ACM, pp. 389–391.

[35] REICHELT, S., HÄUSSLER, R., FÜTTERER, G., AND LEISTER, N. Depth cues in human visual perception and their realization in 3d displays. In *Proc. SPIE* (2010), vol. 7690, p. 76900B.

[36] SAVINO, P. J., AND DANESH-MEYER, H. V. *Color Atlas and Synopsis of Clinical Ophthalmology – Wills Eye Institute – Neuro-Ophthalmology (Wills Eye Institute Atlas Series)*. Lippincott Williams & Wilkins, 2012.

[37] SIBERT, L. E., AND JACOB, R. J. K. Evaluation of eye gaze interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2000), CHI '00, ACM, pp. 281–288.

[38] STELLMACH, S., AND DACHSELT, R. Look &#38; touch: Gaze-supported target acquisition. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2012), CHI '12, ACM, pp. 2981–2990.

[39] SUNDSTEDT, V. Gazing at games: using eye tracking to control virtual characters. In *ACM SIGGRAPH 2010 Courses* (2010), ACM, p. 5.

[40] TURNER, J., VELLOSO, E., GELLERSEN, H., AND SUNDSTEDT, V. Eyeplay: applications for gaze in games. In *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play* (2014), ACM, pp. 465–468.

[41] Velloso, E., Carter, M., Newn, J., Esteves, A., Clarke, C., and Gellersen, H. Motion correlation: Selecting objects by matching their movement. *ACM Trans. Comput.-Hum. Interact. 24*, 3 (Apr. 2017), 22:1–22:35.

[42] Velloso, E., Wirth, M., Weichel, C., Esteves, A., and Gellersen, H. Ambigaze: Direct control of ambient devices by gaze. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems* (2016), ACM, pp. 812–817.

[43] Vertegaal, R., Mamuji, A., Sohn, C., and Cheng, D. Media eyepliances: using eye tracking for remote control focus selection of appliances. In *CHI'05 Extended Abstracts on Human Factors in Computing Systems* (2005), ACM, pp. 1861–1864.

[44] Vidal, M., Bulling, A., and Gellersen, H. Detection of smooth pursuits using eye movement shape features. In *Proceedings of the symposium on eye tracking research and applications* (2012), ACM, pp. 177–180.

[45] Vidal, M., Bulling, A., and Gellersen, H. Pursuits: spontaneous interaction with displays based on smooth pursuit eye movement and moving targets. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing* (2013), ACM, pp. 439–448.

[46] Vidal, M., Pfeuffer, K., Bulling, A., and Gellersen, H. W. Pursuits: eye-based interaction with moving targets. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems* (2013), ACM, pp. 3147–3150.

[47] Špakov, O., Isokoski, P., Kangas, J., Akkil, D., and Majaranta, P. Pursuitadjuster: An exploration into the design space of smooth pursuit –based widgets. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications* (New York, NY, USA, 2016), ETRA '16, ACM, pp. 287–290.

[48] Ware, C., and Mikaelian, H. H. An evaluation of an eye tracker as a device for computer input2. In *ACM Sigchi Bulletin* (1987), vol. 17, ACM, pp. 183–188.

[49] Zhang, Y., Bulling, A., and Gellersen, H. Sideways: A gaze interface for spontaneous interaction with situated displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2013), ACM, pp. 851–860.