UNIVERSITÄT DES SAARLANDES
Department of Computer Science
Media Informatics

MASTER'S THESIS

Over There!
Visual Guidance in 360-Degree Videos and Other Virtual Environments

Author
Christoph Alexander Rosenberg (255597)

September 21, 2017

Supervisor/Additional Reviewer
Prof. Dr. Antonio Krüger
Dr. Michael Schmitz

Advisor
Marco Speicher, M.Sc.

Universität des Saarlandes
Saarland University

Hochschule der Bildenden Künste Saar
Academy of Fine Arts Saar

Declaration of Authorship

Erklärung

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Statement

I hereby confirm that I have written this thesis on my own and that I have not used any other media or materials than the ones referred to in this thesis.

Einverständniserklärung

Ich bin damit einverstanden, dass meine (bestandene) Arbeit in beiden Versionen in die Bibliothek der Informatik aufgenommen und damit veröffentlicht wird.

Declaration of Consent

I agree to make both versions of my thesis (with a passing grade) accessible to the public by having them added to the library of the Computer Science Department.

Datum/Date

Unterschrift/Signature

Abstract

Christoph Alexander Rosenberg

Over There!
Visual Guidance in 360-Degree Videos
and Other Virtual Environments

360-degree videos offer a new viewing experience with the ability to explore virtual environments much more freely than before. However, technologies and aesthetics behind this approach of film-making are not yet fully developed. The newly gained freedom creates challenges as new methods have to be established to guide users through narratives.
This work provides an overview of methods to guide users and gathers information from two user studies that explore visual guidance in 360-degree videos. The studies utilise videos filmed specifically to test the potential of user guidance. The methods are designed to maintain immersion while still providing guidance. Three devices are used to examine possible differences. Methods include effects added in post-production as well as seemingly simple actions like gestures.
Subtle guidance is shown to be an ambitious task that is still in need of improvement. However, several methods offer the potential to guide users without breaking immersion and interfering with the users' freedom.

Contents

Declaration of Authorship **i**

Abstract **ii**

1 Introduction **1**

2 Related Work **2**
- 2.1 360-Degree Videos 2
  - 2.1.1 Production 2
  - 2.1.2 Formats 3
- 2.2 Guidance in Traditional Film 4
- 2.3 Guidance in Virtual Environments 5

3 Methods of User Guidance **6**
- 3.1 Classification 6
- 3.2 Overview of Methods 9
  - 3.2.1 Forced Rotation 9
  - 3.2.2 Object to Follow 10
  - 3.2.3 Person to Follow 11
  - 3.2.4 Object Manipulation 11
  - 3.2.5 Environment Manipulation 12
  - 3.2.6 Gestures 13
  - 3.2.7 No Guidance 13
- 3.3 Additional Methods 15
  - 3.3.1 Sound 15
  - 3.3.2 Subtle Gaze Direction 15
  - 3.3.3 Interface 16
  - 3.3.4 Cuts 17
  - 3.3.5 Accelerated Viewpoint Panning 18

4 Studies **19**
- 4.1 Pre-Study 19
  - 4.1.1 Hypotheses 20
  - 4.1.2 Participants 21
  - 4.1.3 Apparatus 21
  - 4.1.4 Design 22
  - 4.1.5 Task 24

4.1.6 Procedure 24
4.1.7 Results 25
Devices 25
Methods 29
4.2 Main Study 31
4.2.1 Hypotheses 31
4.2.2 Participants 31
4.2.3 Apparatus 31
4.2.4 Design 32
4.2.5 Task 39
4.2.6 Procedure 40
4.2.7 Results 40
Devices 41
Methods 44
Participants 48

5 Discussion **49**
5.1 Pre-Study 49
5.1.1 Devices 49
5.1.2 Methods 50
5.2 Main Study 51
5.2.1 Devices 51
5.2.2 Methods 51
No Guidance 51
Forced Rotation 52
Object to Follow 53
Person to Follow 53
Object Manipulation 53
Environment Manipulation 54
Gestures 54
5.2.3 Additional Results 55

6 Conclusion **57**

7 Future Work **58**

8 Further Reading **59**

A Web Application **60**

B Study Questionnaires **61**
B.1 Pre-Study Questionnaires 61
B.1.1 NASA Task Load Index 61
B.1.2 User Experience Questionnaire 62
B.1.3 Motion Sickness Awareness Questionnaire 63

B.1.4 Immersion Questionnaire 63
B.1.5 Demographic Questionnaire 64
B.2 Main Study Questionnaires 65
B.2.1 Questions After a Video 65
B.2.2 Questions After the Experiment 66

Bibliography **67**

List of Figures

2.1 Equirectangular Projection . . . 3

4.1 Pre-Study Methods . . . 20
4.2 Pre-Study Camera Set-Up . . . 21
4.3 Pre-Study Set-Up . . . 22
4.4 Pre-Study Objects . . . 23
4.5 Pre-Study Participant . . . 24
4.6 Number of Gestures/Devices . . . 26
4.7 Error Rate/Devices . . . 26
4.8 Overall Task Load/Devices . . . 26
4.9 Task Load Ratings/Devices . . . 26
4.10 Overall User Experience/Devices . . . 27
4.11 User Experience Ratings/Devices . . . 27
4.12 Overall Motion Sickness/Devices . . . 28
4.13 Immersion/Devices . . . 28
4.14 Search Time/Methods . . . 29
4.15 Number of Gestures/Methods . . . 29
4.16 Error Rate/Methods . . . 30
4.17 Error Rate/Devices/Methods . . . 30
4.18 Method: No Guidance/Game: Hit the Pot . . . 33
4.19 Method: Forced Rotation/Game: Musical Chair . . . 33
4.20 Method: Object to Follow/Game: Blowing Wool . . . 34
4.21 Method: Person to Follow/Game: Hide-and-Seek . . . 35
4.22 Method: Object Manipulation/Game: Ping Pong . . . 35
4.23 Method: Environment Manipulation/Game: Rubber Ducks Race . . . 36
4.24 Method: Small Gestures/Game: Can Knock-Down . . . 37
4.25 Method: Big Gestures/Game: Throwing Balls . . . 37
4.26 Main Study Participant . . . 39
4.27 Error Rate/Devices . . . 41
4.28 Task Load Performance/Devices . . . 41
4.29 Overall User Experience/Devices . . . 42
4.30 User Experience Ratings/Devices . . . 42
4.31 Overall MSAQ/Devices . . . 43
4.32 Immersion/Devices . . . 43
4.33 Performance/Methods . . . 44

4.34 Performance Accuracy/Methods 45
4.35 Overall Task Load/Methods 45
4.36 Overall User Experience/Methods 46
4.37 User Experience Ratings/Methods 46
4.38 Overall UEQ/Performance Accuracy/Methods 47
4.39 Overall MSAQ/Methods 47
4.40 MSAQ Dizzy/Methods 47
4.41 Experience/Accuracy 48

A.1 Web application 60

List of Abbreviations

HMD	**H**ead-**M**ounted **D**isplay
MSAQ	**M**otion **S**ickness **A**ssessment **Q**uestionnaire
NASA TLX	NASA **T**ask **L**oad **I**ndex
SUS	**S**later **U**soh **S**teed (Immersion Questionnaire)
UEQ	**U**ser **E**xperience **Q**uestionnaire

1 Introduction

In March 2015, YouTube started to support 360-degree videos. In September 2015, Facebook made 360-degree videos available in their news feed. 11 months after the roll-out on Facebook, more than 250,000 360-degree videos had been uploaded to the service. In 2016, nearly 89 million virtual reality headsets were expected to be sold, most of them utilising smartphones. These numbers do not only show a rise of interest in 360-degree videos and virtual reality but also the accessibility of this technology to consumers (Verma, 2015; Saba, 2015; Beddoe-Stephens, 2016a; Korolov, 2016).

Despite this, the technologies and aesthetics behind this approach of film-making are not yet fully developed. Techniques and rules that were established during the history of film cannot be transferred completely to this new format (L. T. Nielsen et al., 2016). Users gained the freedom to explore their virtual surroundings independently from intentions of the film-makers. This introduces a number of problems to both content creators and film makers in all kinds of virtual environments. Guidance of the user during the exploration is a crucial aspect.

This work contributes an overview and categories of methods to guide users in virtual environments and especially in 360-degree videos. Several of these methods were tested in two user studies and aim to provide guidance while maintaining the freedom and immersion the technology offers. These studies create a foundation to be able to further improve and refine the applied techniques.

2 Related Work

2.1 360-Degree Videos

360-degree videos are "full spherical videos" "covering the complete field of view" (F. Nielsen, 2005, p. 92). To view these "immersive spherical videos", they are mapped to a virtual sphere "where the user can look around" (Hosseini and Swaminathan, 2016, p. 1). As virtual reality is defined by Brooks, 1999, p. 16 as an experience "in which the user is effectively immersed in a responsive virtual world", a 360-degree video cannot be a virtual reality as it is not responsive. Nonetheless, a virtual environment is created (Chen, 1995).

2.1.1 Production

Several ways to watch 360-degree videos exist. This includes mobile devices as well as large domes with projected videos (Ramalho and Chambel, 2013). Filming 360-degree videos usually involves a multihead camera design. These cameras use more than one lens and image sensor to cover the full field of view. However, this set-up induces several problems as multiple video streams have to be combined into a single image (F. Nielsen, 2005; Philpot et al., 2017).
This process is called stitching. It is similar to the process used to combine still images to a panorama. Overlapping areas of the videos make it possible to create a single surround video by computing the necessary distortions. In addition to the stitching, parallax errors can occur because of the different focal planes of the sensors. Distortions in the video are visible when objects are close to the sensors and cross the seams between the cameras in the video (F. Nielsen, 2005).

FIGURE 2.1: Equirectangular projection of a 360-degree video

2.1.2 Formats

360-degree videos are saved as two-dimensional videos utilising a variety of mapping techniques. These include spherical and cubical, which project the video on a sphere or the sides of cube respectively. However, the equirectangular projection is commonly used which projects the video onto a rectangular plane with a 2:1 aspect ratio, covering 360 degrees horizontally and 180 degrees vertically. To do this, the image has to be distorted, similar to cartography when mapping the Earth on a plane surface (F. Nielsen, 2005).

Sound in 360-degree videos can utilise the same standard formats as traditional videos (YouTube, 2017). Still, different approaches exist. Spatial audio uses several techniques to give users the possibility to experience sounds dynamically in the virtual environment. This aims to improve immersion and lets users locate sound sources. To do so, characteristics of the human way of hearing are incorporated (Google, 2016).

Another approach is directional audio. User's choose what they want to hear in the scene by directing their gaze in that direction. Audio sources that are not in front of the user are then muted. This technique aims to let users experience more than one story at once by utilising the freedom of virtual environments (Google, 2017).

2.2 Guidance in Traditional Film

Traditional film uses several techniques to guide attention which were established during the history of film. While technology evolved, these techniques did also change and were adapted to new formats such as colour or sound (L. T. Nielsen et al., 2016). However, virtual environments require different approaches as established methods may not be transferable.

Cinematography is one way to guide attention by defining how the image is filmed. This includes camera settings like focal length or aperture. Using these settings, effects like depth of field can be created: the important part of the image remains in focus while the rest of the scene is blurred. Furthermore, camera movement can be deployed to show certain parts of the scene or move in on details (L. T. Nielsen et al., 2016; Danieau, Guillo, and Doré, 2017; Ascher and Pincus, 2012). As 360-degree videos show the whole scene by definition, most of these techniques cannot be applied.

Everything in front of the camera can be changed. This can be lighting and of course the narrative itself. Certain features in the picture may specifically attract attention. This can be colour, luminance or motion. This can also be used in 360-degree videos, although it may be perceived differently and it might not always be possible to adjust the scene to guide the attention accordingly (Danieau, Guillo, and Doré, 2017; Ascher and Pincus, 2012).

The freedom to look around enables the user to frame own images. This makes it impossible for the creator of the film to stay in control of the narrative as the user is not forced to follow it (L. T. Nielsen et al., 2016; Danieau, Guillo, and Doré, 2017). While editing is used to control the narrative in traditional film, this is limited in 360-degree videos. Cuts imply a change of location and are therefore prone to motion sickness and lower immersion (Bozgeyikli et al., 2016).

As Hata, Koike, and Sato, 2016 note, guidance should be as subtle as possible to not interrupt the viewers' concentration or create annoyance.

Sound is another way to guide the attention in films. Users are attracted by sounds and try to localise it (Johnson and Proctor, 2004, p. 105). This is possible in virtual environments as well. However, as the environment is three-dimensional, this may be harder to achieve.

2.3 Guidance in Virtual Environments

L. T. Nielsen et al., 2016 introduces three options to present a narrative in virtual reality:

1. The narrative does only continue if the user has perceived certain elements of the scene. It is halted otherwise.
2. The system takes care of introducing the key elements of the directly to the user.
3. Cues are used to guide the attention and steer the user to the area of interest.

These points show the different approach to storytelling in virtual environments as compared to traditional media. However, 360-degree videos are not interactive as virtual reality is. These approaches, especially to halt the narrative, may be harder achieve in videos than in content that is rendered in real time. The methods introduced in this study focus on the third point.

3 Methods of User Guidance

A wide variety of methods to guide users through virtual environments exist. The featured methods are described in the context of 360-degree videos. However, most methods and characteristics can also be used in different environments including rendered content like virtual reality. As these environments are more interactive than prerecorded videos, alternative or additional methods may be feasible but are not covered here (see Future Work, chapter 7). Furthermore, in virtual environments that require the user to move around, methods to physically reorient the user in the real world can have similar approaches as methods guiding the user's gaze (Suma et al., 2012).

3.1 Classification

A number of properties is used to classify the different methods. The advantages and disadvantages of the methods are reflected in these four main categories:

Performance The performance is the most important characteristic of each method. It determines the success in guiding the user. A method that guides the user successfully to the area of interest does therefore have a good performance. If the performance is not satisfying, the other characteristics will usually not outweigh this disadvantage. It is therefore the key characteristic.
While the performance of several methods was measured in the studies, it is subject to change depending on the exact implementation. Further information on the performance of these methods is available in Studies, chapter 4.

Parameters The parameters describe the necessary circumstances of the environment and the scene through which the user is guided.
The potential size of the area of interest varies depending on the method. If the method allows to point to a small area of interest, a better guidance is expected. Guidance to specific details in the scene instead of merely the approximate direction may provide a better performance.
To let the user understand and subsequently follow the guidance, some methods may need to be established by introducing the method before the actual guidance occurs. This could cause problems if the method is used without context and not easily understandable when used in a single, short instance (e.g. Object to Follow, subsection 3.2.2). If guidance is possible in those circumstances, the possibility of isolated use is given.
Some methods may require a certain context in the scene. This includes the presence of persons (or similar characters) to perform gestures or other tasks (e.g. Person to Follow,

subsection 3.2.3). The context may be essential and could render the method pointless if not available as the user would not be able to understand the guidance.
The naturalness defines whether the techniques employed by a method are a realistic part of the scene or on a meta level. While gestures are usually a natural part of the scene (e.g. Gestures, subsection 3.2.6), methods with more extensive effects are not (e.g. Environment Manipulation, subsection 3.2.5). This affects the immersion, but can also impact the performance of the method (Danieau, Guillo, and Doré, 2017, p. 206).

User Experience To classify the experience of the user during the guidance, several aspects are examined.
Immersion is one of the most important goals in virtual reality:

> The experience of being transported to an elaborately simulated place is pleasurable in itself, regardless of the fantasy content. We refer to this experience as immersion. Immersion is a metaphorical term derived from the physical experience of being submerged in water. We seek the same feeling from a psychologically immersive experience that we do from a plunge in the ocean or swimming pool: the sensation of being surrounded by a completely other reality, as different as water is from air, that takes over all of our attention, our whole perceptual apparatus. (Murray, 1997, p. 98)

To maintain the immersion, the methods to guide the user should be subtle (Suma et al., 2012). To achieve this, the concept of the method should not challenge the user so that the attention remains on the actual content. Guiding the user without affecting immersion is one of the main goals of these methods. The ease of use describes how difficult it is to grasp this concept and follow the guidance. This is also relevant considering that additional workload in virtual environments can cause confusion and frustration (Van den Broeck, Kawsar, and Schöning, 2017).
A method has to be perceptible as it couldn't guide the user if it were unnoticeable. However, it is also desirable to be as subtle as possible as more perceptibility could potentially have an impact on the immersion (Suma et al., 2012).
An important aspect of virtual environments is motion sickness. Symptoms like disorientation or nausea may occur and some methods could contribute to this (e.g. Forced Rotation, subsection 3.2.1) as they fit the common theory of "a sensory mismatch or conflict between the vestibular and visual senses" (Davis, Nesbitt, and Nalivaiko, 2014).

Production The implementation of the methods in 360-degree videos requires a varying amount of work. Depending on the method, this can occur during all stages of the production. While some effects can be added during post-production to arbitrary videos (e.g. Danieau, Guillo, and Doré, 2017), other methods require specific planning before shooting (e.g. Person to Follow, subsection 3.2.3). This can include the movement of persons in the scene which can usually not be changed easily in post-production.
The addition of effects in post-production does require less planning as changes to the

video can be reversed or altered. However, the workload for the implementation is also depending on the method and may vary for each scene even if the same method is used (e.g. Object to Follow, subsection 3.2.2). To keep the methods subtle, the amount of visual differences between the original video and the version with methods added should be minimal but still perceivable.
Some methods may also impose the necessity for special equipment. This can be during production (e.g. Sound, subsection 3.3.1) or during playback (e.g. Forced Rotation, subsection 3.2.1).
These categories can be extended from 360-degree videos to virtual reality. While most methods are transferable, the implementation or certain aspects may have to be adjusted.

3.2 Overview of Methods

This overview presents a range of methods to guide the user. These methods and their performance are covered in the two studies and additional data is available in Studies, chapter 4.

3.2.1 Forced Rotation

Forced rotation moves the area of interest into the user's viewport by rotating the virtual environment. While the user is free to look around in general, the system takes control during the rotation. This behaviour mirrors that of traditional films as the area of interest is not chosen by the user but predetermined (Danieau, Guillo, and Doré, 2017; L. T. Nielsen et al., 2016).
A notable implementation of forced rotation was introduced by Facebook in 2016. 360-degree videos uploaded to Facebook can be edited to include points of interest and the camera will pan to these points at a specified time during playback. The feature can be disabled by the user (Beddoe-Stephens, 2016b).

Parameters The method solely manipulates the direction of the user's viewport. Therefore, the size of the area of interest matches the size of the viewport and is not defined any further. There are no requirements regarding the content of the scene as the method does not interact with it. Isolated use of forced rotation is possible but may confuse the user and is not a natural part of the video in most cases (Danieau, Guillo, and Doré, 2017; L. T. Nielsen et al., 2016).

User Experience No active interaction is required from the user, so forced rotation is easy to learn. However, users may misunderstand the rotation of the camera which could create confusion and frustration. This can happen when the rotation is misinterpreted as camera movement and the user tries to counterbalance this by rotating the video in the opposite direction (Danieau, Guillo, and Doré, 2017).
The discrepancy between the movement of the camera in the virtual environment and the physical movement of the user renders forced rotation prone to motion sickness. Additionally, the method is very perceivable (Danieau, Guillo, and Doré, 2017).

Production Forced rotation does not require planning and the method can be added during post-production. No elaborate effects are used and the rotation is the only difference between the original and the edited video.
When used in 360-degree videos, there are two possible implementations. The rotation can be pre-rendered and therefore be a part of the actual video or it could be done by the video player. In Facebook's implementation, the video player handles the rotation. The points of interest are defined and the video is rotated accordingly. As information about the user's gaze is available during playback, the rotation can be adjusted dynamically. This requires support by the video player and additional information about the points

of interest not stored in the video file (Beddoe-Stephens, 2016b).
The second implementation pre-renders the video with the rotation. No additional information about points of interest is provided, so the approximate direction of the user's gaze during playback has to be predicted in the editing process. This implementation could cause further confusion when the actual gaze of the user during playback differs but does not need support by the video player. As virtual reality cannot be pre-rendered, this constraint does not exist in that environment.

3.2.2 Object to Follow

This method uses an object to guide the user through the scene. An example of such an object is the firefly used by L. T. Nielsen et al., 2016. While the user is able to look around the environment independently, the firefly in the scene "offered clues as to where the user should focus" (L. T. Nielsen et al., 2016, p. 230).
A similar approach was used to reorient the user by Peck, Fuchs, and Whitton, 2009. In this case, a sphere distracted the user from manipulations of the virtual environment. This includes the guidance of the user's gaze, albeit using this for a different purpose (reorientation). Improved versions of this method replaced the sphere with a butterfly and a hummingbird as these are "more natural for the [virtual environment] being used" (Peck, Fuchs, and Whitton, 2009, p. 6).

Parameters The object can guide the user to small areas of interest by flying near them. As the user has to grasp the concept of the method, use of the method in isolated instances is not possible without initially establishing the object as a guide.
The object itself does not require any context to work. However, to use the method as a natural manipulation of the scene, the context of the scene has to be suitable (e.g. using animals in nature).

User Experience To establish the object as a guide in the scene and have the user follow it, the user has to understand the purpose of the object. This could be achieved by letting it "hoover in one place when relevant information [is] presented in that area of the scene" (L. T. Nielsen et al., 2016, p. 230) and doing the same when the area of interest changes even when following the narrative is not a problem so that the user is able to get used to the concept.
The object is perceivable by design. This is necessary but affects immersion. Even a natural integration can be identified as a guide and could disrupt the experience. This is also evident by the use of similar objects as distractors in the experiments by Peck, Fuchs, and Whitton, 2009. Motion sickness should not differ from the original video.

Production The object that is used by the method is the only visible difference from the original video. It is added during post-production where the amount of work is moderate. No specific planning is necessary in earlier stages of production and no special equipment is required.

The implementation of this method in virtual reality might be easier than the visual effect in a video. In this case, the virtual environment most likely already uses a number of objects and an additional one to guide the user could be added easily.

3.2.3 Person to Follow

This method uses a person or a similar character to guide the user through the scene. It is an approach to integrate the guide in the scene instead of adding an object as in Object to Follow, subsection 3.2.2. This integration marks the main difference between these two methods.

Parameters The area of interest a guide may refer to can be very small. The guide can also use gestures and facial expressions. Depending on the scene, this method can be very natural as the guide is a part of the narrative. However, the guide has to be established as an important element so the user can follow the movement. This requires a context in which a guide can exist. A scene without a person is not compatible with this method and additionally, the guide has to be able to move around.

User Experience The person acting as a guide is perceivable. If an appropriate context is given and the guide is established, an impact on immersion or motion sickness is unlikely.

Production No post-production is necessary and there are no visual differences as the guide is already in the scene. The implementation may still be very complex as a lot of planning is required to integrate the guide in the scene seamlessly and film it accordingly.
In virtual reality, the structure of the scene has to be planned as well. Creating the guide for this environment may lead to further work.

3.2.4 Object Manipulation

Object manipulation directs the user's gaze by manipulating objects in the virtual environment and gaining the user's attention through these changes to the scene. Manipulations can include movement, changes in colour, contrast, luminance and shape or a combination of these and other aspects (Mateescu and Bajić, 2014). Areas containing "unpredictable contours or unusual details" (Mackworth and Morandi, 1967) are known to attract the user's attention.

Parameters The manipulation of an object in the scene draws attention to this object in particular. The area of interest is therefore defined by the object itself and its immediate surroundings. To integrate the method naturally in the scene, there has to be a context for the manipulation, e.g. a television displaying images on its screen. Isolated use is possible as the method does not need to be established.

User Experience Gaining the attention of the user by object manipulation is a natural process and does not require the user to become accustomed to the method. The manipulations are perceptible while the concept of the method may not be obvious to the user if the changes are a natural part of the scene. Based on this, the only suspected impact on immersion occurs if the manipulations are not properly integrated.
Motion sickness is not expected to be influenced.

Production Depending on the particular kind of the manipulation, planning and execution of the method require additional effort. To integrate the manipulations naturally, specific planning is needed especially in regard of the positioning of the manipulations so that they are visible to the user. The visual changes to the original video vary but are limited to the manipulated object. Some manipulations could be executed on set instead of being edited in afterwards.
This method works very similar in virtual reality.

3.2.5 Environment Manipulation

In contrast to object manipulation (see Object Manipulation, subsection 3.2.4), the manipulation of the environment as used in this method manipulates everything but the area of interest in the scene. Danieau, Guillo, and Doré, 2017 and Hata, Koike, and Sato, 2016 describe different manipulations to guide the user to this unedited part of the video: Fade to black, desaturation and blur. Other manipulations like a change of contrast or a combination of these effects are feasible (see Main Study, section 4.2).
All these effects can vary in intensity and may be animated to guide the user more precisely by progressing to the area of interest instead of appearing in every part of the scene at the same time. They are all based on the principle that the attention focusses on the area where the effect is not applied and the image remains unaltered. The intensity of the effects may progressively increase depending on the distance to the area of interest. The effects are only applied for a short amount of time during which the user is guided.
The concept of this method is derived from techniques used in traditional film like depth of field and lighting (see Related Work, chapter 2). Guidance of the user is one of the purposes of these techniques and is adapted here (Ascher and Pincus, 2012).

Parameters As almost the whole scene is manipulated, the area of interest has to be identifiable by the user and should therefore not be too small. The method does not require a certain context and can be used in isolated instances. Applying the described effects to the whole image is a very unnatural manipulation.

User Experience The user has to be able to perceive the changes in the video and understand it in order to focus on a certain part of the scene. This can cause confusion and frustration as the intent of the manipulations may be unclear (Danieau, Guillo, and Doré, 2017).

As the effects in the video have to be clearly visible, immersion will be affected. Due to the large areas of the scene being manipulated, motion sickness may occur. This mirrors the effects of these techniques when used in traditional film: "[I]f something important is out of focus, the viewer may feel annoyed or uncomfortable" (Ascher and Pincus, 2012, Chapter 4).

Production Most of the video is altered by the effects, creating large differences to the original video during these parts. The implementation has to be done in post-production. Even with an animation, these effects are usually relatively simple to implement and require no specific planning. Furthermore, lighting the set during filming could be an alternative to adding effects in post-production.
The effects can be applied in virtual reality as well.

3.2.6 Gestures

This method uses gestures by persons or similar characters who point the user to the area of interest. This covers small gestures, e.g. a nod, and big gestures, e.g. pointing in a direction. An even more subtle approach are facial expressions (see Pre-Study, section 4.1).

Parameters Gestures point the user in a general direction without being able to refer to small details. They are very natural and can be used without being established (Raza Abidi, Williams, and Johnston, 2013). However, at least one person has to be in the scene to make the gesture.

User Experience As "[h]umans point to direct attention and influence behaviour" (Raza Abidi, Williams, and Johnston, 2013, p. 67), gestures are easy to understand. They are perceptible but natural and are not expected to have an impact on immersion or motion sickness.

Production The gestures are recorded during the filming of the scene so that some planning is necessary. This includes the timing and type of the gestures. Post-production is not necessary.
In virtual reality, gestures can also be implemented.

3.2.7 No Guidance

Despite not actually guiding the user, the absence of a method may also provide advantages. However, the performance is expected to be inferior to all actual methods.

Parameters No explicit technique is used to guide the user. There are no restrictions in terms of context, isolated use, naturalness of manipulations or the potential size of the area of interest.

User Experience The user does not need to learn anything and nothing is perceptible or could create motion sickness. Despite this, confusion and frustration may arise if following the narrative of the video becomes more difficult as the user is not guided. This could also impact immersion which is otherwise not different from the original video. However, these effects are theoretical.

Production As there is nothing added or altered in the video, no planning or editing is required.

3.3 Additional Methods

This section introduces additional methods which are not covered in the two studies (see Studies, chapter 4) for several different reasons (see the descriptions for more details).

3.3.1 Sound

Auditory cues are a very important element of attention guidance. Sounds evoke the urge to localise their source and "can have strong effects on the orienting of visual attention" (Johnson and Proctor, 2004, p. 105). This can be utilised to gain the attention of the user for guidance to the area of interest.

Parameters The size of the area of interest varies depending on the audio system and the sounds. If a very specific sound gains the attention of the user and there is only one possible location of origin (e.g. the sound of a door with only one door present in the virtual environment) the source is obvious to the user. Other sounds may be harder to locate especially if no surround sound is used. In the right context within the scene, sounds are very natural.

User Experience Due to the natural occurrence of the sound, immersion should be improved by using sounds. It is very easy for the user to understand the concept and there is no impact on motion sickness.

Production As there are different ways of producing sounds and using them in 360-degree videos, the amount of planning and the work load for the implementation varies. Examples for this are spatial and directional audio (see Formats, subsection 2.1.2). There may be a need for microphones on set or surround sound editing in post-production. These restrictions do not apply in virtual reality where this can be calculated in real time during usage.

Despite the great potential of sound in user guidance, it is not covered in-depth in the studies as this work focusses on visual guidance methods. Additional equipment and inordinate effort would have been necessary to cover sound and its possible variations in attention guidance.

3.3.2 Subtle Gaze Direction

Subtle gaze direction manipulates areas in the peripheral gaze of the user to direct the user's attention to a certain part of an image. Manipulations include modulations of luminance or colour. These are terminated before the user's gaze reaches these areas which is possible due to the usage of eye-tracking technology (Bailey et al., 2009). Subtle gaze direction is capable of supporting and guiding the user in more complex scenarios (McNamara, Bailey, and Grimm, 2008) and in narrative art (McNamara, Booth, et al., 2012). The technique has also been adapted for videos (Sridharan and Bailey, 2016) and virtual environments (Sridharan, Pieszala, and Bailey, 2015).

Parameters The viewer is directed to an area of interest, however, as the manipulation is terminated before the user's gaze reaches this area, the user cannot locate the exact location and therefore, the area of interest has to be sized appropriately (Bailey et al., 2009). The manipulations are unnatural, but remain barely noticeable under ideal conditions. No specific context is required and isolated use is possible.

User Experience The method is not directly perceivable by the user, however, the perceived image quality may be lowered (Bailey et al., 2009). Similarly, immersion may be affected. An influence on motion sickness is unlikely.
The user does not need to understand the method, as following the manipulations is a natural, instinctive process. Frustration and confusion should remain minimal, although this may change as user's response to the stimuli differs (Bailey et al., 2009).

Production While the amount of visual differences is minimal, planning may be required to determine the positions of the manipulations. This creates additional work during post-production especially as the method is not yet completely implemented or thoroughly tested in virtual environments. Furthermore, the method requires eye-tracking technology which creates further challenges and makes playback far more complicated (Bailey et al., 2009; Sridharan, Pieszala, and Bailey, 2015).

Subtle gaze direction may have potential as a method to guide users, however, there are still a number of restrictions preventing practical use. This includes eye-tracking technology which is not widely available, especially in mobile devices and head-mounted displays. Additionally, a specialised playback system is necessary to incorporate the manipulations into the virtual environment and dynamically adjust these in respect to the eye-tracking data. Subtle gaze direction is not part of the studies.

3.3.3 Interface

Several approaches to guide users in virtual environments with the help of interfaces exist. These include maps, radars and arrows (Chittaro and Burigat, 2004) as well as specialized techniques designed specifically for virtual environments (Wonner et al., 2013). Depending on the method, the interfaces indicate the user's position in the virtual space or show the route or direction to the area of interest. A prominent implementation of an interface in 360-degree videos is Facebook's heading indicator (Beddoe-Stephens, 2016a). Other approaches, including text overlays and interactive interfaces, have been discussed (Neng and Chambel, 2010).

Parameters These interfaces can provide guidance to small areas of interest and can be used without context with the trade-off of usually being an unnatural part in 360-degree videos.

User Experience Despite being intuitive for most users if well designed, an interface has to be perceivable and will have an impact on immersion in most cases, as it can rarely be a natural part of the scene. Motion sickness and confusion are unlikely to emerge.

Production Visual differences are clearly visible in the final video. No additional planning is necessary, but the areas of interest have to be defined and during playback, the interface has to be rendered by the player. This is done by Facebook in combination with their forced rotation technique (see subsection 3.2.1) where a heading indicator shows the direction of the user's gaze (Beddoe-Stephens, 2016b).
Due to the interactive nature of virtual reality, interfaces may be easier to implement in virtual environments of this type and if an interface for other purposes is already in use (e.g. in games with a head-up display), it might feel more natural.

Due to the studies being focused on subtle approaches and interfaces usually not being part of videos, this method is not included in the studies.

3.3.4 Cuts

Despite being one of the most important elements in traditional film-making, editing is often much more limited in 360-degree videos (see Related Work, chapter 2) (Ascher and Pincus, 2012). However, cuts can still be used in 360-degree videos and guidance of the user may be one reason for doing so.

Parameters Cuts do not require a special context and can be used in isolated instances. In contrast to traditional film, the area of interest cannot be cut to appear exactly in front of the user as the rest of the scene is still visible in the virtual environment and the user is free to look around. Although cuts are an established part of films, they might be less natural in 360-degree videos.

User Experience In a virtual environment, cuts are very perceptible and unnatural, so an impact on immersion is to be expected. This can also lead to confusion and motion sickness, as sudden interruptions like cuts do not appear in real-world environments. This could also lead to difficulties for the user to understand what is happening in the scene.

Production While different perspectives and editing are an ubiquitous part of traditional film, the same is harder to achieve in virtual environments. Camera positions have to be changed and planned accordingly as a static 360-degree camera cannot provide techniques such as specific framing of an image or zoom.

While cuts and other forms of editing are usually expected in a traditional film, 360-degree videos work differently. This does also mean that these techniques have to be used in other ways. Editing is vital for storytelling and may also be used in virtual environments. However, as a method solely used to direct the attention in a scene without further purpose, cuts may not be an appropriate technique. Because of this, editing is not included as a method in the studies.
In virtual reality, other kinds of transitions may be more pertinent. This does include user controlled actions like teleportation (see Further Reading, chapter 8) (Bozgeyikli et al., 2016).

3.3.5 Accelerated Viewpoint Panning

Accelerated viewpoint panning is not a method to guide the user's attention directly. It is designed to make it easier to look around in virtual environments by accelerating the movements of the user (Hong and Kim, 2016). This constant or dynamic gain in panning is unnoticeable to the user and should facilitate the process of gathering information in the virtual environment. It can be applied in addition to other effects and in virtual reality environments.

4 Studies

Two studies have been conducted to gather information about different methods of user guidance and their performance on different devices. The main study builds upon the findings of the pre-study to further refine the implementations of some of the methods. Both studies feature several videos to be watched by the participants to identify objects they were guided to.
While sound offers huge potential as a method of user guidance in combination with its several possible implementations, the following studies focus on visual methods. These were selected to subtly guide users through 360-degree videos with a wide range of techniques. However, methods that require special equipment (subtle gaze control) or were deemed unfitting for subtle guidance in videos (interface, cuts) were not included. Videos from the studies are attached to the physical copies of this thesis and available online at https://tiny.cc/OverThere.

4.1 Pre-Study

The pre-study was planned to explore the differences of three types of devices when watching 360-degree videos. Additionally, three types of gestures were tested to determine their eligibility as guidance methods. Users were asked to identify objects they were pointed to by the gestures in the videos on the different devices.
The device were selected to represent those available to consumers. They were classified in three categories:

- Desktop
  This category consists of desktop computers and notebooks. Users usually sit while using these devices and look around in the virtual environments with input from the mouse.

- Mobile
  Smartphones and tablets are part of this category. Navigation in virtual environments involves physical movement of the devices, although input with gestures (e.g. rotating videos by swiping across the screen) may be possible as well.

- Head-Mounted Displays
  Head-mounted displays expect the user to physically look around while wearing the device. This includes advanced headsets like the HTC Vive or Oculus Rift as well as other head-mounted solutions (Samsung Gear VR or Google Cardboard).

FIGURE 4.1: The three methods used in the pre-study: facial expression, small gesture, big gesture

The three types of gestures used in the videos consisted of:

- Facial Expressions
  The facial expressions used in this study consisted mostly of eye movement. Slight movement of the head was necessary in some instances when the placement of the objects in the scene prevented direct eye contact.
- Small Gestures
  Small gestures were defined as head movements towards the object in addition to the facial expressions. This also included leaning in the direction of the guidance.
- Big Gestures
  In addition to the facial expressions and small gestures, the big gestures also featured pointing to the object.

4.1.1 Hypotheses

H_1 Head-mounted displays offer the best performance and immersion followed by mobile devices with desktop devices performing worst.

H_2 Head-mounted displays are preferred by the users.

H_3 Head-mounted displays are most prone to motion sickness and mobile devices more prone to motion sickness than desktop devices.

H_4 The performance of big gestures is better than that of small gestures while facial expressions have the worst performance on all devices.

4.1.2 Participants

A total of 18 participants took part in the pre-study. They were aged between 22 and 32 years (*M* = 25.39, *SD* = 2.453) and consisted of 12 males and 6 females. 77.8 % of the participants had already watched 360-degree videos on at least one device. 72.2 % were familiar with this kind of content on a desktop computer or notebook, the same number on a smartphone or tablet and 61.1 % on a head-mounted display.
To ask participants if their previous experience was positive, a Likert scale from 1 to 5 was used (1: Strongly disagree; 5: Strongly agree). While all participants who had previously watched 360-degree videos on a head-mounted display agreed or strongly agreed to have had a positive experience, 61.5 % did so with regard to their experience on mobile devices and also 61.5 % on desktop devices.
While none of the participants suffered from colour-blindness, half of them were in need of optical aids which could be worn during the experiments.

4.1.3 Apparatus

FIGURE 4.2: The camera set-up used in the pre-study

The videos used for the study were recorded on a Samsung Gear 360 camera in a resolution of 3840 x 1920 with 29.97 frames per second. The final stitched and edited videos used in the study had a resolution of 4096 x 2048 pixels and no sound was included. The videos were rendered in the equirectangular format.
The desktop set-up consisted of a standard desktop computer and a 24 inch display with a resolution of 1920 x 1080 pixels and 60 Hz. A standard optical mouse was used as the

input device for the participants which required to click and drag the picture to navigate. All videos were played in the GoPro VR Player by Kolor on Microsoft Windows 10.
Mobile devices were represented by an LG Nexus 5X smartphone featuring a 5.2 inch display with a resolution of 1080 x 1920 pixels (60 Hz) and Android 6.0. VRTV VR Video Player Free by Chai Software was used to play the videos with a sphere projection. Users were able to look around in the virtual environment by moving the device.
An HTC Vive was used to represent head-mounted displays, featuring a resolution of 2160 × 1200 pixels (1080 × 1200 per eye) and 90 Hz. The videos were provided by the same desktop computer and software as above with the exception of the mouse input which was not used.
The field of view was adjusted on all devices to approximately 100 degrees.

4.1.4 Design

FIGURE 4.3: The set-up of the pre-study in equirectangular format from a video used in the study

The study used a total of 30 360-degree videos plus an additional introduction video. Every video featured the same room from the same static camera position. In the videos, a woman is standing in the room approximately one meter from the camera. She faced the camera and used the defined gestures to point to an object placed in the room. There was a total of ten objects with one object being pointed to in each video. The objects were placed around the room to allow gestures in all directions. The gestures were repeated five times within each video resulting in videos with a length of about 35 seconds. This amounted to 3 gestures x 10 objects = 30 videos.

FIGURE 4.4: A detailed view of the objects used in the pre-study

These were the ten objects featured in the videos including their unique details the participants had to state in addition to the object itself:

1. Drink (brand: Pepsi)
2. Poster (character: Homer Simpson)
3. Shirt (colour: green)
4. Backpack (colour: blue)
5. Toy Figure (character: Minion)
6. Waste Bin (colour: yellow)
7. Book (colour: yellow)
8. Film (title: The Simpsons Movie)
9. Bananas (quantity: two)
10. Candy (brand: Riesen)

Every participant watched a total of 30 videos on a single device. These videos were divided into three blocks with each block consisting of ten videos. All videos in one block used the same gesture. The order of the videos in each block was randomized while the order of the three blocks was counterbalanced using Latin squares.
As each participant watched the videos on a single device, the three devices were used by six participants each. This amounted to 3 devices x 6 participants x 3 gestures/blocks x 10 objects/videos = 540 trials.
This resulted in two independent variables:

- Device (desktop, mobile, HMD)
- Gesture (facial expression, small gestures, big gesture)

The dependent variables were defined as:

- Performance (task completion time in seconds and the number of gestures necessary, identification of the object, identification of the object detail)
- Preference (task load, user experience, motion sickness, immersion/presence)

There was no time limit, but each video could only be watched once. Participants took about 50 minutes to complete the experiment, including the questionnaires.

4.1.5 Task

Every participant had to perform a single task for each video they watched: identification of the object that was indicated. When an object was identified, the video was stopped immediately and the time in seconds and number of gestures up to this point in the video were recorded. After the video was stopped, participants were asked to identify the object and the unique detail.

FIGURE 4.5: A participant watches a 360-degree video on the desktop device during the pre-study

4.1.6 Procedure

After the participant was welcomed to the study, an introduction video was shown to practice the controls of the device. During the study, all participants remained seated on a swivel chair. Each participant received information on the handling of the device and was free to ask questions. The introduction video featured the room with all objects and the woman present, however, she did not perform any gestures.
Following this, the participant watched the 30 videos and performed the aforementioned task for each. At the start of each video, the participants faced the same direction

both in the real world and in the virtual environment at the start of each video. In the video, they were facing the woman in the room. As soon as the video had started, they were free to look around. They were neither forced nor instructed to watch the woman or her gestures.
Finally, when all 30 videos had been watched, the participant filled out five questionnaires (see Study Questionnaires, section B.1):

1. NASA Task Load Index (TLX) (So, 2017)
2. User Experience Questionnaire (UEQ) (Laugwitz, Held, and Schrepp, 2008)
3. Motion Sickness Assessment Questionnaire (MSAQ) (J. Gianaros et al., 2001)
4. Immersion/Presence Questionnaire (SUS) (Slater, Usoh, and Steed, 1994)
5. Demographic Questionnaire

4.1.7 Results

The results of the pre-study are covered in this section. Analyses of variance were used to find significances.

Devices

This section covers the results of the pre-study regarding the three different types of devices used in the pre-study.

Performance The search time is defined as the time between the first gesture in a video and the answer of the participant. It represents the time needed by the participant to determine which object was referenced by the gestures. The number of gestures indicates how many gestures were shown in the video until the answer of the participant. As the gestures are repeated five times in each video, five is the maximum number of gestures while the maximum search time is the length of the video minus twelve seconds (the first gesture in each video starts after twelve seconds).
The mean search time on the desktop computer was 7.07 seconds (SD = 5.666), on the mobile device 7.68 seconds (SD = 5.380) and on the head-mounted display 8.46 seconds (SD = 6.341). These differences were not significant ($p > 0.05$). Similarly, the mean number of notifications until an answer was given was 2.00 (SD = 1.260) on the desktop computer, 2.18 (SD = 1.372) on the mobile device and 2.42 (SD = 1.429) on the head-mounted display. Here, a significant effect was found ($p < 0.05$) (see Figure 4.6).
The error rate represents how often the wrong or no object was identified. While the lowest search times and notification numbers were recorded on the desktop device, the error rate was the highest (26.7 %). On the mobile device and the head mounted display the error rate was lower at 17.8 % and 19.4 % respectively (see Figure 4.7).
The details of the objects were not perceived by 30 % of the participants on the desktop devices. Once more, this error rate was lower on the other devices with 21.7 % on both.

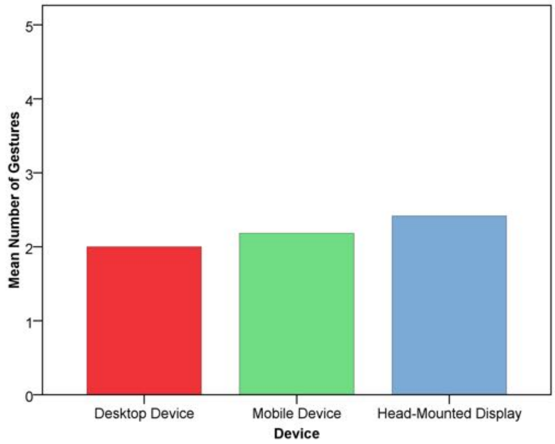

FIGURE 4.6: Number of Gestures/Devices

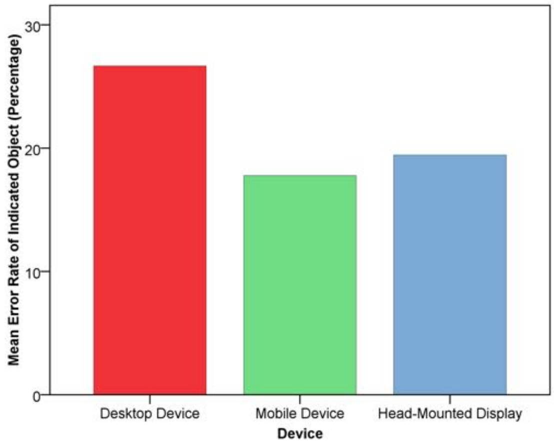

FIGURE 4.7: Error Rate/ Devices

Task Load The NASA Task Load Index was used in the pre-study to measure the participants' task load on a scale from 0 to 100. The questionnaire was filled out once per participant, so the results describe the task load of the device used and not that of the different gestures.

The overall task load is lowest on the head-mounted display (*M* = 35.61, *SD* = 22.162) and highest on the mobile device (*M* = 51.94, *SD* = 15.693) with the desktop device in between (*M* = 40.00, *SD* = 15.287) (see Figure 4.8).

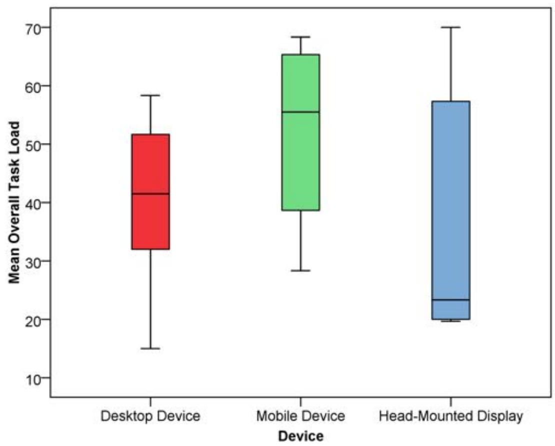

FIGURE 4.8: Overall Task Load/Devices

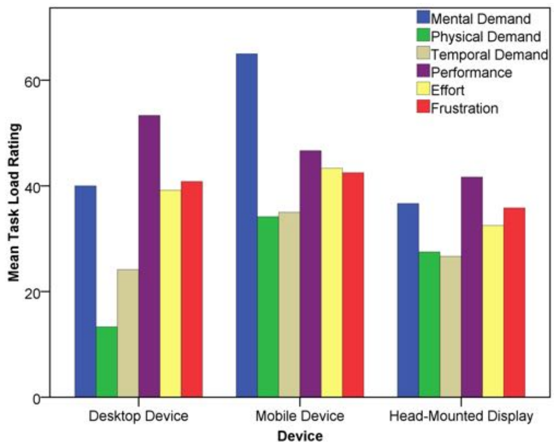

FIGURE 4.9: Task Load Ratings/Devices

While the mental demand was similar on the desktop device and the head-mounted display (*M* = 40.00, *SD* = 23.238 and *M* = 36.67, *SD* = 23.594), it was higher on the mobile device (*M* = 65.00, *SD* = 14.832), albeit not significantly ($p > 0.05$). The physical demand was lower on the desktop computer (*M* = 13.33, *SD* = 11.690) than on both the mobile phone and the head-mounted display (*M* = 34.17, *SD* = 28.003 and *M* = 27.50, *SD* =

12.550). The participants rated their performance the highest on the desktop device, followed by the smartphone with the head-mounted display last. No significances were found (see Figure 4.9).

User Experience The User Experience Questionnaire rates the user experience in six categories on a scale from -3 to 3. Values higher than 0.8 represent a positive evaluation. Overall, comparing the mean of all categories, the mobile device was rated the lowest (*M* = 0.48, *SD* = 0.734), followed by the head-mounted display (*M* = 0.84, *SD* = 0.837) with the desktop computer rated best (*M* = 1,18, *SD* = 0.281) (see Figure 4.10). The individual categories were rated in the same order by the participants with no significances (see Figure 4.11).

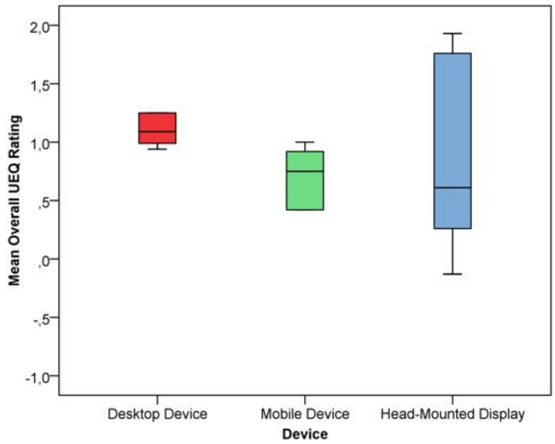

FIGURE 4.10: Overall UEQ/ Devices

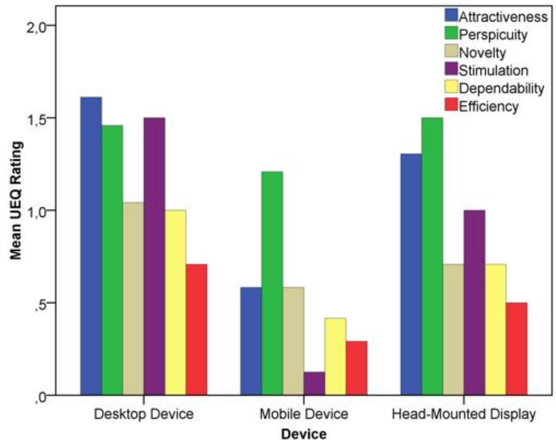

FIGURE 4.11: User Experience Ratings/Devices

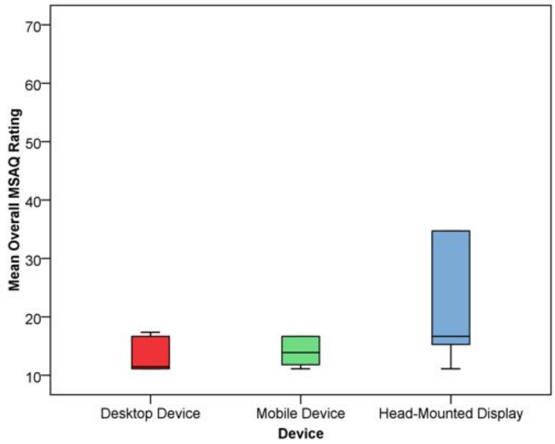

FIGURE 4.12: Overall MSAQ/Devices

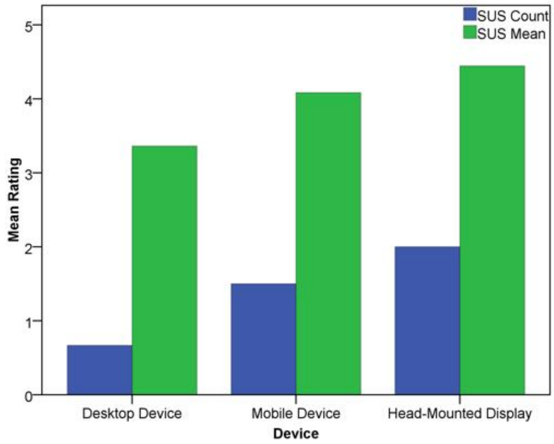

FIGURE 4.13: Immersion/Devices

Motion Sickness The Motion Sickness Assessment Questionnaire rates the motion sickness of the participants on a scale from 0 to 100 (percentage of maximum number of points). Overall, the head-mounted display was most prone to motion sickness (M = 27.08, SD = 21.677). The mobile device and the desktop computer follow with a lower score (M = 16.09, SD = 6.753 and M = 13.19, SD = 2.979) (see Figure 4.12).

Immersion To measure the immersion and presence of the virtual environment, the questionnaire from Slater, Usoh, and Steed, 1994 was used. Participants answered six questions on a Likert scale from 1-7 with the results represented by the SUS count (Slater Usoh Steed) and the SUS mean. The SUS count represents the number of answers with a high rating (i.e. 6 or 7 on the Likert scale; maximum count of 6) while the SUS mean represents the mean of all answers (scale from 1 to 7).
SUS count and SUS mean showed similar results with no significance: The head-mounted display was rated the highest (SUS count: M = 2.00, SD = 1.265; SUS mean: M = 4.44, SD = 0.993), the mobile device having lower scores (SUS count: M = 1.50, SD = 0.837; SUS mean: M = 4.08, SD = 1.079) and the lowest scores on the desktop computer (SUS count: M = 0.67, SD = 0.816; SUS mean: M = 3.36, SD = 1.335) (see Figure 4.13).

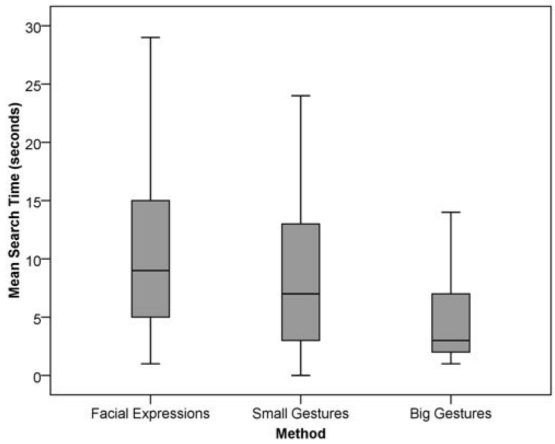

FIGURE 4.14: Search Time/Methods

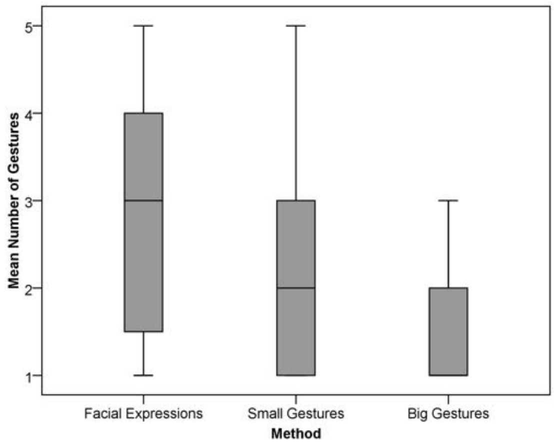

FIGURE 4.15: Number of Gestures/Methods

Methods

In the pre-study, three types of gestures were tested including facial expressions.

Performance When facial expressions were used, it took participants a mean of 9.66 seconds (*SD* = 5.958) and 2.82 notifications (*SD* = 1.454) to determine which object was indicated. Using small gestures, it were 8.53 seconds (*SD* = 6.236) and 2.22 gestures (*SD* = 1.313) and 5.01 seconds (*SD* = 4.585) and 1.56 gestures (*SD* = 0.981) using big gestures. The differences were significant ($p < 0.01$) (see Figure 4.14 and Figure 4.15).
40.6 % named either no object or not the one indicated when using facial expressions. The error rate was highest for objects placed behind the woman making the gestures with 40.3 %. The total error rate was 18.3 % with small gestures and 5.0 % with big gestures. The unique details of the objects were not perceived by 43.3 %, 20.6 % and 9.4 % respectively. These differences were significant ($p < 0.01$) (see Figure 4.16).
On the desktop device, small gestures lead to an error rate of 33.3 % while it was only 11.7 % and 10.0 % on the mobile device and the head mounted display respectively. While facial expressions did also lead to the worst error rate on the desktop device (43.3 %; mobile: 36.7%; HMD: 41.7 %), big gestures had the lowest error rate on this device (3.3 %; mobile: 5.0 %; HMD: 6.7 %) (see Figure 4.17).

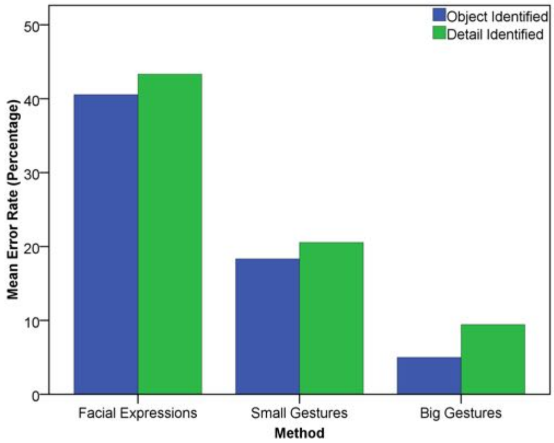

FIGURE 4.16: Error Rate/Methods

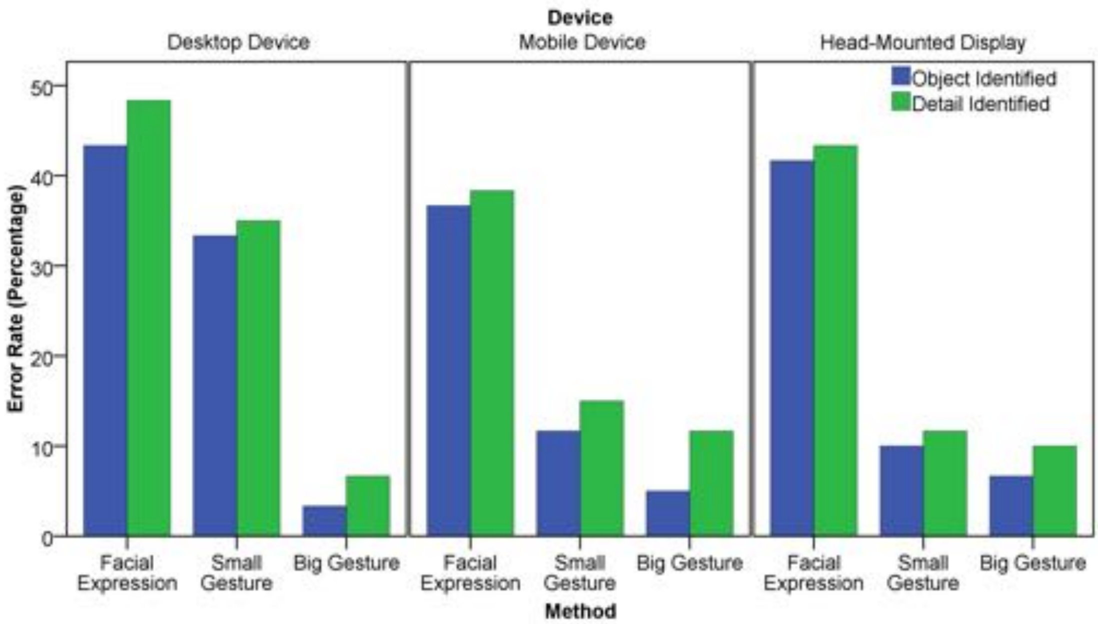

FIGURE 4.17: Error Rate/Devices/Methods

4.2 Main Study

The purpose of the main study was to find out more about the different methods for visual guidance. The design of the study was built upon the design and the findings of the pre-study (see Pre-Study, section 4.1).

4.2.1 Hypotheses

H_0 Visual guidance improves the performance.

H_1 Methods that create less differences to the original video rate higher in terms of immersion.

H_2 Unsuccessful guidance impairs the user experience.

H_3 Head-mounted displays outperform mobile devices in all aspects (performance and preference).

4.2.2 Participants

The main study included 32 participants who consisted of 18 males and 14 females aged between 22 and 29 years (M = 25.19, SD = 1.942). 37.5 % of the participants had no or almost no experience with 360-degree videos or virtual reality. 46.9 % had some experience and 15.6 % were very experienced.
Out of the 71.9 % of participants who were experienced with 360-degree videos on head-mounted displays, only 65.2 % rated their experience as positive (mobile: 62.5 %; desktop: 57.9 %). The frequency of watching 360-degree videos was rated on a Likert scale from 1 to 5 (1: Never; 5: Very often). 93.8 % of the participants never or only rarely watched 360-degree videos (1 or 2 on the Likert scale; M = 1.59, SD = 0.712).
None of the participants suffered from colour-blindness, but 59.4 % needed some kind of optical aid which could be worn during the experiments.

4.2.3 Apparatus

The same camera and video formats as in the pre-study (see Pre-Study Apparatus, subsection 4.1.3) were used. Videos were filmed with a Samsung Gear 360 in 3840 x 1920 pixels with 29.97 frames per second and the final videos again used 4096 x 2048 pixels with the same frame rate and no sound in the equirectangular format.
As a mobile device, an LG Nexus 5X was used as in the pre-study (5.2 inch display with a resolution of 1080 x 1920 pixels and 60 Hz). During this study, the smartphone was updated to Android 7.1 and VR Player FREE by SpherePlay was used to play the videos. These changes were not noticeable by the participants and did not affect the videos.
The HTC Vive was used again with a resolution of 2160 × 1200 pixels (1080 × 1200 per eye) and 90 Hz. A standard desktop computer provided the videos utilising the Kolor

GoPro VR Player on Windows 10. The field of view remained unchained as well at approximately 100 degrees.

4.2.4 Design

Considering the results from the pre-study, the device types were reduced to mobile and head-mounted display. The desktop computer is not used in the main study. The desktop device had the highest error rate in the pre-study. It was decided to concentrate on the other two devices to get better results regarding the different methods of guidance. Another factor was the higher immersion rating of the head-mounted display and the smartphone which was deemed an important characteristic.
The main study utilised a total of 79 360-degree videos, including an introduction video. The same room as in the pre-study was featured, again from a static camera position. In the videos, two men played a variety of party games and a method for user guidance was used in each video to direct the user's attention to a manipulated object.
The setting as a birthday party with several games was chosen to be able to watch the videos in arbitrary order as one game was featured in each video. Additionally, the games worked without sound, enabling the possibility of using the same videos for all participants despite the experiment and the questionnaires being available in English and in German. The games served as a narrative to provide a simple, easy to understand story in the short videos (35 - 60 seconds).
Eight different games were recorded:

- Hit the Pot
- Musical Chair
- Hide-and-Seek
- Blowing Wool
- Ping Pong
- Rubber Ducks Race
- Can Knock-Down
- Throwing Balls

Each video contained an object that was manipulated to change its colour for three seconds during the video. During these three seconds, the colour pulsated three times to make the manipulation more obvious. The duration served as a time limit during which the participants had to react to the guidance and prevented that the manipulation could be seen by chance at another point in the video without guidance. One of the eight methods chosen for the study was then implemented to guide the user to the object.

No Guidance No guidance was used as a reference. The manipulations are present in the videos but the user is not guided towards them (see Figure 4.18).

FIGURE 4.18: Method: No Guidance/Game: Hit the Pot

Forced Rotation The manipulation is rotated into the gaze of the user and the video is rotated back once the manipulation has ended. Depending on the position of the manipulation, the rotation itself lasted between one and three seconds to maintain approximately the same velocity over all combinations of games and objects.
The rotations were part of the video and not calculated by the player. They were based on the probable location of the user's gaze in the scene (i.e. the game in the scene) and were added during post-production (see Figure 4.19).

FIGURE 4.19: Method: Forced Rotation/Game: Musical Chair

Object to Follow This method was implemented as a combination of the methods used by Peck, Fuchs, and Whitton, 2009 and L. T. Nielsen et al., 2016. The object to follow was a bright, yellow sphere which was flying close to the areas of interest. Its movements were animated to resemble that of a firefly. It hovered close to the game in the scene, moved to the manipulation when it occurred and flew back after it concluded (see Figure 4.20).

FIGURE 4.20: Method: Object to Follow/Game: Blowing Wool

Person to Follow The person to follow was part of the game hide-and-seek. A person searches the other person's hiding place around the room or is looking for a place to hide. The manipulated object appears in the proximity of the person going around the room so it is visible when the user is following the search (see Figure 4.21).

FIGURE 4.21: Method: Person to Follow/Game: Hide-and-Seek

Object Manipulation To implement the object manipulation, the already existing manipulations of pulsating colours were altered by adding an additional glow effect. This effect increased the size of the manipulation and was modelled after the glow of a television set in a dim environment (see Figure 4.22).

FIGURE 4.22: Method: Object Manipulation/Game: Ping Pong

Environment Manipulation Several of the possible manipulations were combined for the environment manipulation. This includes changes in brightness and saturation as well as blur. However, the implementation differed from the ones used by Danieau, Guillo, and Doré, 2017 and Hata, Koike, and Sato, 2016. In addition to the combination of the effects, the manipulations were also animated to gradually increase their size in the scene until only the area of interest remained unaltered. Because of this, the user was supposed to be able to follow the manipulations to the area of interest. Therefore, the manipulations were clearly noticeable (see Figure 4.23).

FIGURE 4.23: Method: Environment Manipulation/Game: Rubber Ducks Race

Small Gestures The small gestures introduced in the pre-study were used. These include looking and nodding in the direction of the area of interest two times. The number of repetitions was reduced to generate a more realistic scenario. In the videos, this was done by one person to show the other person an object in the room. This object or an object in the close proximity of it was manipulated (see Figure 4.24).

FIGURE 4.24: Method: Small Gestures/Game: Can Knock-Down

Big Gestures The big gesture of pointing directly in the direction of the area of interest introduced in the pre-study was implemented again. This was done in the same way as the small gestures (see Figure 4.25).

FIGURE 4.25: Method: Big Gestures/Game: Throwing Balls

As facial expressions performed worse than the other methods in the pre-study, especially regarding search time and error rate, this method was omitted from the main study. This decision was also motivated by the fact that each gesture was only repeated two times instead of five and the person executing the gestures not facing the camera directly.
The gestures and the person to follow were methods that had to be implemented during the filming of the scene. The other methods were added in post-production. Therefore, all of the games except hide-and-seek were used for all methods that could be added during post-production, leading to five versions of each game. Hide-and-seek was used exclusively for the method person to follow. The small and big gestures were implemented in three games each (blowing wool, can knock-down, throwing balls) by filming the games again with the gestures incorporated into the scene (e.g. pointing in the directions of the balls).
Every game featured a unique combination of manipulated object and colour. Examples of the manipulated objects in the room are a tile in the ceiling, the door of a cabinet, a waste bin, the screen of a television and a window. The manipulations occurred at different times in the videos and were placed around the room.
Additionally, two takes of each game were shot. Every user watched the videos on a smartphone and a head-mounted display to experience all methods on both devices. As two versions of the games were filmed, the content differed slightly despite the games being the same. The manipulated object was different in each take. This amounted to 2 takes x 7 games x 5 methods = 70 videos, with 8 additional videos featuring the person to follow and gestures (2 takes of hide-and-seek/person to follow and 2 takes/gestures * 3 games). The introduction video featuring the arrival of one person in the room with a birthday cake to establish the story increased the number of videos to 79.
The independent variables were similar to those in the pre-study:

- Device (mobile, HMD)
- Method (eight methods)

The dependent variables were defined as:

- Performance (detection of guidance, identification of the manipulated object, identification of the method)
- Preference (task load, user experience, motion sickness, immersion/presence)

Every participant watched a total of 16 videos. No video was watched twice. Eight videos were watched on each device featuring all of the games with all of the methods in a counterbalanced order. This was possible as two takes of each game were recorded. While all participants experienced all methods and all games twice (on different devices), the exact combination of these varied.
In total, this amounted to 2 devices/takes x 8 games/methods x 32 participants = 512 trials.

4.2.5 Task

After each video was watched completely, participants were asked if something had been indicated to them. In the case of a positive answer, they were asked to identify the indicated object and the method of guidance used in the video. If they stated that nothing had been indicated, these two questions were omitted.

FIGURE 4.26: A participant watches 360-degree videos on the smartphone and the head-mounted display during the main study

4.2.6 Procedure

The participants were welcomed to the experiment and had the opportunity to become accustomed to 360-degree videos by watching the introduction video. After that, the first eight videos were shown. The participants remained seated on a swivel chair during the experiment. No specific task was given, however, after the completion of each video, participants were asked if something had been indicated to them and if that happened, what was indicated and how it was done. In addition to these questions, additional questionnaires had to be answered about each video regarding the user experience, motion sickness, presence/immersion and task load. The questionnaires were based on the same questionnaires as in the pre-study. To be able to give answers after every video to gather data about every method on every device from every participant, these questionnaires were shortened and customized (see Study Questionnaires, Appendix B).

When the participants finished watching the first eight videos, the device was changed and the same procedure was repeated on the second device. When all 16 videos had been watched, a final demographic questionnaire was filled out.

4.2.7 Results

As the User Experience Questionnaire, the Motion Sickness Assessment Questionnaire and the immersion questionnaire from Slater, Usoh, and Steed, 1994 were shortened (see Study Questionnaires, Appendix B), the results are not directly comparable to other studies including the pre-study. The NASA Task Load Index remained unchanged.

The User Experience Questionnaire consisted of six questions that represented each category of the original questionnaire (Likert scale from 1 to 7). The overall result was calculated accordingly with a final score on a scale from -3 to 3 and the individual questions were adjusted to this scale as well. The same was done with the Motion Sickness Assessment Questionnaire with one question from each of the four original categories. The final score was again the percentage of the maximum number of points. The immersion questionnaire was reduced to a single question with a comment section. Users rated their sense of being in the room on a Likert scale from 1 to 7.

In the performance results, no equivalent of the search time from the pre-study exists. This is due to the participants watching each video completely and the manipulations only occurring for three seconds. The maximum search time is therefore three seconds, however, to be realistically able to perceive the manipulation it has to be even lower.

Significances were found using analyses of variance.

Devices

The results of the two different devices used in the main study (smartphone and head-mounted display) are covered in this section.

Performance 68.8 % of the participants did not name the correct object that was indicated in the videos using the mobile device. This error rate was 61.7 % on the head-mounted display. The method of guidance was not correctly recognised by 59.8 % when watching the videos on the smartphone. Using the head-mounted display, this error rate was 54.3 % (see Figure 4.27). The differences were not significant ($p > 0.05$).

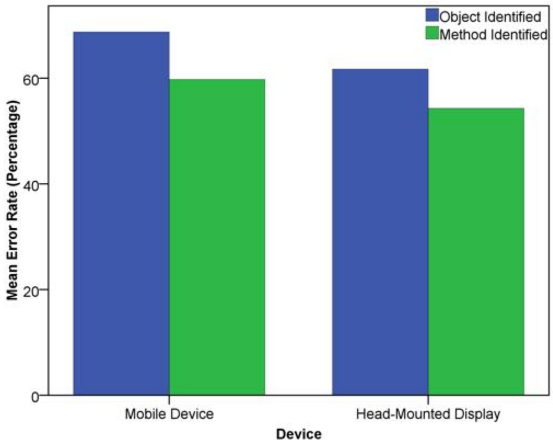

FIGURE 4.27: Error Rate/Devices

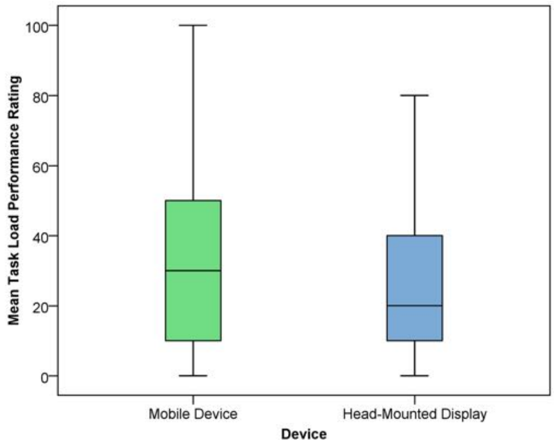

FIGURE 4.28: Task Load Performance/Devices

Task Load The overall task load was rated similar on both devices (mobile: $M = 30.23$, $SD = 18.746$; HMD: $M = 29.29$, $SD = 19.658$). However, there was a significant effect ($p < 0.05$) found in the performance rating: Participants rated their own performance better on the mobile device ($M = 31.52$, $SD = 22.361$) than on the head-mounted display ($M = 27.15$, $SD = 23.399$) (Figure 4.28).

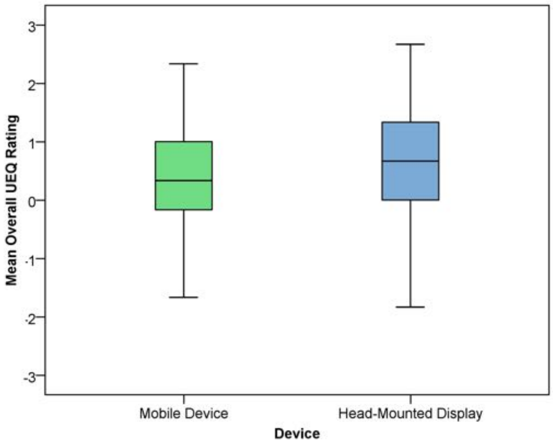

FIGURE 4.29: Overall UEQ/Devices

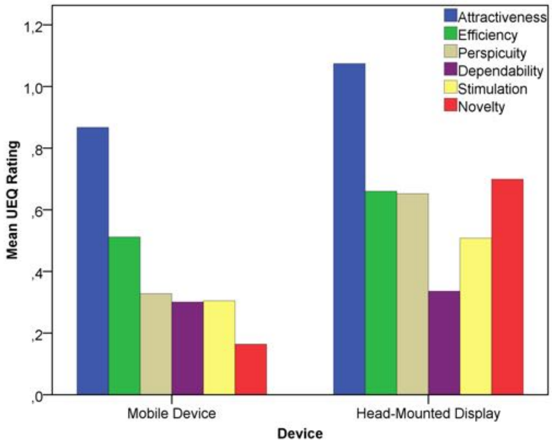

FIGURE 4.30: User Experience Ratings/Devices

User Experience The overall score for the user experience showed significant differences between both devices ($p < 0.01$): The mobile device had a value of 0.41 ($SD = 0.838$) and the head-mounted display 0.65 ($SD = 0.939$) (see Figure 4.29). Both values are below the threshold of 0.8 for a positive evaluation (as defined in the full UEQ).
Another significant difference was found in the novelty rating (conventional versus inventive; $p < 0.01$): The smartphone was rated 0.16 ($SD = 1.497$) and the HTC Vive 0.70 ($SD = 1.577$). Significant effects were found in the ratings for perspicuity (confusing versus clear) and stimulation (inferior versus valuable): Perspicuity ($p < 0.05$) was rated 0.33 ($SD = 1.604$) on the mobile device and 0.65 ($SD = 1.640$) on the head-mounted display with stimulation ($p < 0.05$) rated 0.30 ($SD = 1.124$) on the mobile device and 0.51 ($SD = 1.106$) on the head-mounted display (see Figure 4.30).

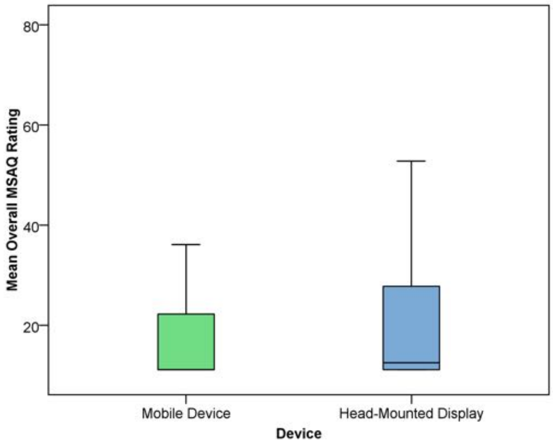

FIGURE 4.31: Overall MSAQ/Devices

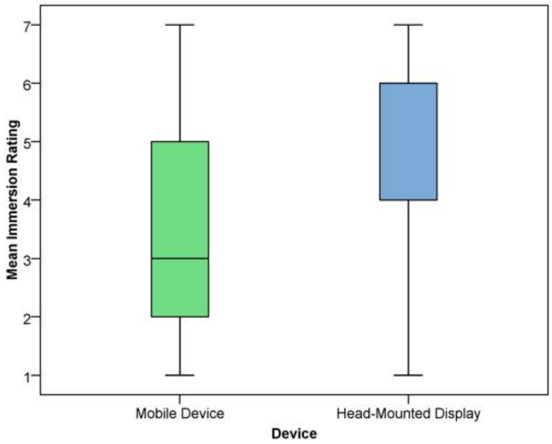

FIGURE 4.32: Immersion/Devices

Motion Sickness The modified Motion Sickness Assessment Questionnaire showed a significant difference in the overall rating ($p < 0.01$): The head-mounted display ($M = 21.53$, $SD = 15.740$) was more prone to motion sickness than the mobile device ($M = 17.97$, $SD = 10.813$). The same was true for the individual questions (see Figure 4.31).

Immersion The immersion was rated on a Likert scale from 1 to 7. There was a significant difference between the devices ($p < 0.01$) with the mobile device rated at 3.59 ($SD = 1.903$) and the head-mounted display at 5.15 ($SD = 1.603$) (see Figure 4.32).

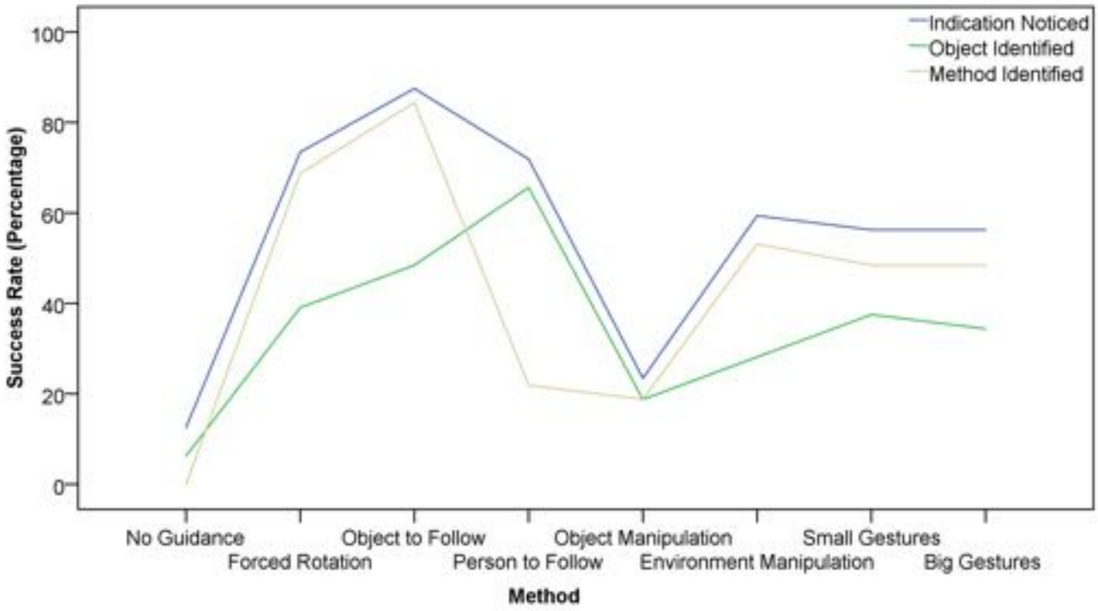

FIGURE 4.33: Performance/Methods

Methods

The results of the eight different methods including the videos with no guidance are covered in this section.

Performance Figure 4.33 and Figure 4.34 summarise the results of the main study in terms of performance. The participants were asked if something had been indicated to them and if their answer was positive, they were asked to name the manipulated object and state the method that guided them there. The differences in the results for these questions were significant ($p < 0.01$).

The accuracy is a combination of the three questions the participants were asked. It assigns a rating depending on what the users identified (nothing; only an indication; an indication and the method; an indication and the object; indication, object and method) (see Figure 4.34).

When no method to guide the user was used, 12.5 % of the participants stated that something had been indicated to them. Half of them (6.3 %) identified the manipulated object correctly. As no method was used, none of the participants could specify the method used for guidance.

Using forced rotation, 73.4 % noted that something was indicated with 68.8 % identifying the method correctly. A total 39.1 % could name the manipulated object.

The method object to follow lead to 87.5 % of the participants stating that something had been indicated and a total of 84.4 % could also name the sphere as the method to guide them. 48.4 % were able to name the manipulated object.

71.9 % of the participants noticed an indication when the method person to follow was

used. 65.6 % were also able to name the manipulation with 21.9 % correctly identified the person in the videos as the method of guidance. This means 48.4 % of all participants were able to name the manipulated object but did not identify any kind of guidance.
In the videos with the manipulation of an object as a method, an indication was perceived by 23.4 %. 18.8 % could name the object and did also identify the method correctly.
Environment manipulation was an indication to 59.4 % of the participants with 53.1 % specifying the manipulation correctly. 28.1 % were able to name the manipulated object.
In the videos with gestures (small or big), 56.3 % noticed an indication and 48.4 % could also name the gestures as the method used. Small gestures led to 37.5 % of the manipulated objects being identified while 34.4 % were able to do so with the big gestures.

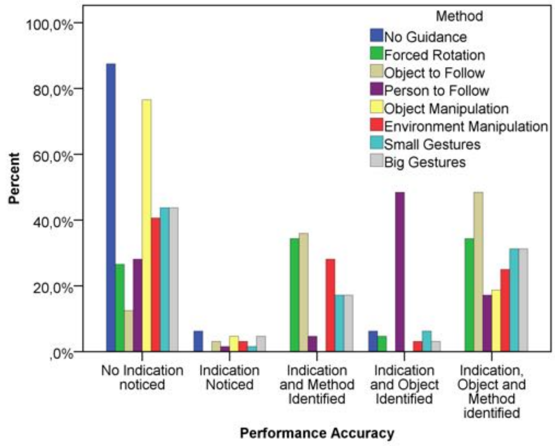

FIGURE 4.34: Accuracy/Methods

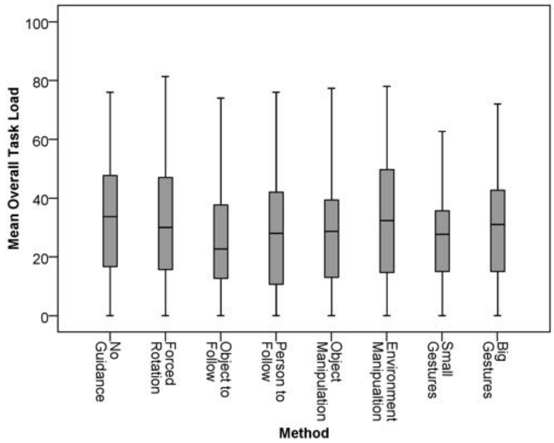

FIGURE 4.35: Overall Task Load/Methods

Task Load The Task Load Index did not yield any significant results. The overall task load was rated between 29 and 33 for most methods with the exception of the methods person to follow(M = 28.4, SD = 19.511), small gestures (M = 27.49, SD = 17.072) and object to follow (M = 25.65, SD = 16.926) (see Figure 4.35).

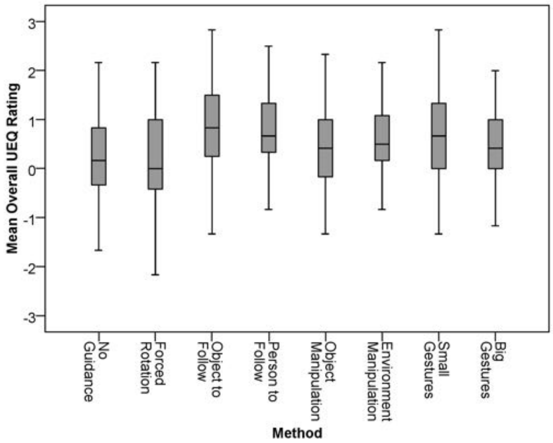

FIGURE 4.36: Overall UEQ/Methods

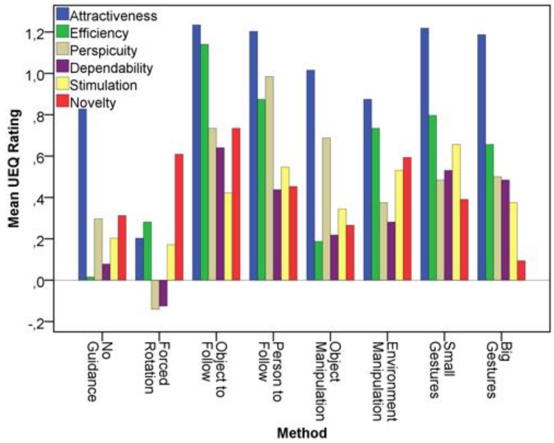

FIGURE 4.37: User Experience Ratings/Methods

User Experience In the overall rating of the user experience (see Figure 4.36), using no guidance or forced rotation led to the lowest ratings (no guidance: $M = 0.29$, $SD = 0.883$; forced rotation: $M = 0.17$, $SD = 1.069$). Person to follow and object to follow were rated highest (person to follow: $M = 0.75$, $SD = 0.839$; object to follow: $M = 0.82$, $SD = 0.920$). The difference was significant ($p < 0.01$).

In attractiveness (unpleasant versus pleasant), forced rotation got the worst result ($M = 0.20$, $SD = 1.575$) with no guidance and environment manipulation being the only other methods rated below 1 (no guidance: $M = 0.83$, $SD = 1.454$; environment manipulation: $M = 0.88$, $SD = 1.254$) (see Figure 4.37).

No guidance, forced rotation and object manipulation were rated least efficient (no guidance: $M = 0.02$, $SD = 1.253$; forced rotation: $M = 0.28$, $SD = 1.496$; object manipulation: $M = 0.19$, $SD = 1.308$).

Forced rotation got the only negative values of these results in perspicuity (confusing versus clear; $M = -0.14$, $SD = 1.735$) and dependability (obstructive versus supportive; $M = -0.13$, $SD = 1.475$).

Big gestures were rated worst in novelty (conventional versus inventive; $M = 0.09$, $SD = 1.477$) with all methods scoring less than 0.8 (threshold for positive evaluation).

When compared to the mean of the accuracy of rating of each method, the overall user experience showed similar trends with the exception of forced rotation.

Motion Sickness While all other methods were rated between 18 and 21 (percentage of the maximum number of points), forced rotation was rated at 23.44 ($SD = 14.872$) (see Figure 4.39). However, the only significant difference ($p < 0.01$) was found when participants rated if they felt disoriented or dizzy (see Figure 4.40): While most methods were rated between 1.88 and 2.11, environment manipulation was rated at 2.2 ($SD = 1.738$) and forced rotation the highest at 2.95 ($SD = 2.171$).

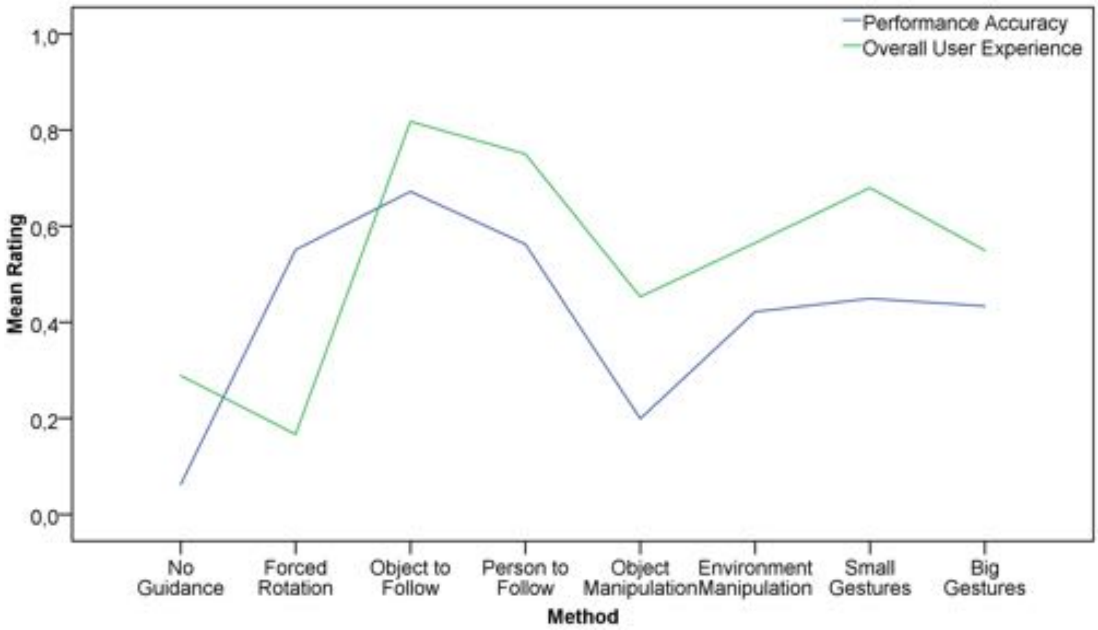

FIGURE 4.38: Overall UEQ/Performance Accuracy/Methods

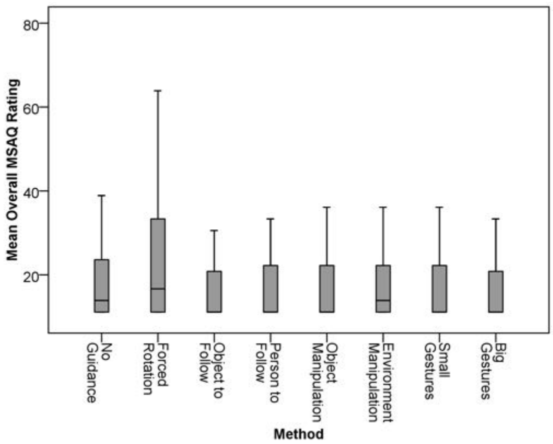

FIGURE 4.39: Overall MSAQ/Methods

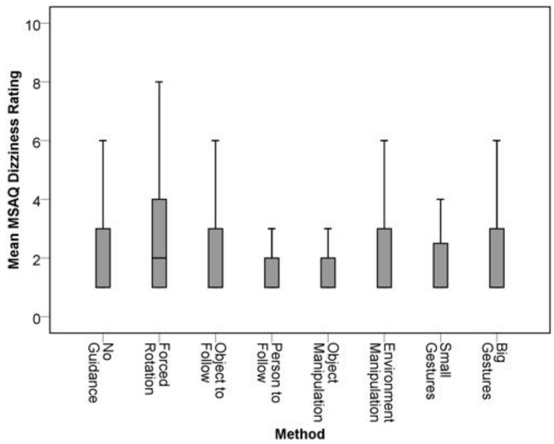

FIGURE 4.40: MSAQ Dizzy/Methods

Immersion The sense of being in the room did not change significantly with the different methods. However, the ratings with no guidance(M = 4.22, SD = 1.812), forced rotation (M = 4.23, SD = 1.998) and environment manipulation (M = 4.25, SD = 2.008) were slightly lower than those of the other methods (ratings between 4.36 and 4.56).

Participants

Results regarding the different participants are covered in this section.

Experience While 49.0 % of the participants without experience with 360-degree videos stated that nothing was indicated to them, it were only 31.3 % of users with extensive experience. 20.8 % of the less experienced users named the manipulated object and the corresponding method correctly. The experienced users did name these correctly in 36.3 % of the videos. These numbers include the videos with no guidance and had an significant effect ($p < 0.05$) (see Figure 4.41).

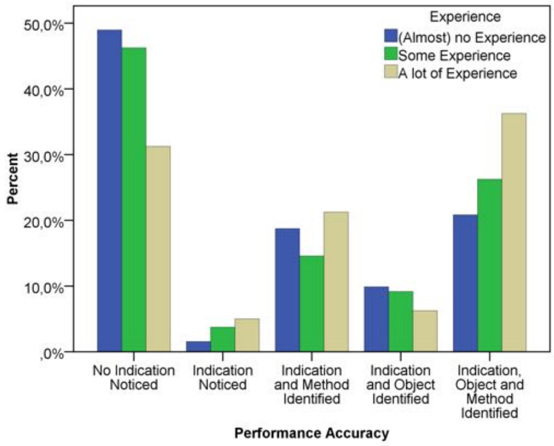

FIGURE 4.41: Experience/Accuracy

Preferences 43.8 % of the participants stated that they prefer traditional videos while 37.5 % prefer 360-degree videos. The rest (18.8 %) has no clear preference and rates the content of the videos as crucial.

When asked if the guidance of attention in videos is annoying, 37.5 % agreed. Nevertheless, the same number of participants disagreed. 25.0 % stated that this depends on method used for guidance.

5 Discussion

5.1 Pre-Study

The pre-study had the goal of gathering data about the three different categories of devices and the three gesture-based methods. This data was to be used to refine the design of the main study.

5.1.1 Devices

In the pre-study, the head-mounted display had the highest search times and participants needed the most notifications to identify an indicated object. The error rate was worse than that of the mobile device. This disproves H_1 partially as the head-mounted display does not outperform the other devices. The time participants needed to determine the object indicated in the video might be longer because more physical movement was necessary. The mouse input of the desktop device solely required movement of one hand, the mobile device required rotation of the swivel chair with the head-mounted display adding the movement of the head to look around. Despite this, the physical demand was rated higher when using the mobile device.

While it took longer to find the object, the error rate was lower on the head-mounted display compared to the desktop computer. This could be caused by a better orientation of the user in the virtual environment thus making it easier for the user to follow the gestures to the indicated object. As the mobile device had an even lower error rate, users might have been able to look around and rotate faster using a smartphone instead of a head-mounted display and therefore being able to better match the gesture of the person to the object while looking around. Additionally, this does further explain the extended search time on the head-mounted display as it is easy to look around very fast when using a mouse but apparently difficult to orientate oneself in the virtual environment. However, the overall task load and the physical demand of the mobile device were rated the highest.

H_1 further states the immersion to be best on the head-mounted display. This is confirmed by the results and as expected, the mobile device performs better in this category than the desktop device. While computer monitors usually offer a much bigger screen size, the experience on mobile devices is improved due to the ability to freely look around physically and this feeling is surpassed by the immersion on the head-mounted display.

H_2 states that users prefer head-mounted displays over the other devices. The overall

task load was indeed the lowest of the three devices. On the one hand the physical demand was higher than that of the desktop computer and the immersion was rated best as expected. On the other hand motion sickness was most noticeable. The desktop computer was rated best in all categories of the user experience. This does not lead to a clear verdict which device users prefer as the results are inconclusive. As each participant did watch all videos on a single device, the data for each device was gathered by a low number of participants. This was improved in the main study.
Head-mounted displays are more prone to motion sickness than mobile devices and desktop computers. This result confirms the expectations in H_3. The differences might have been even more evident if the participants had been standing while using the mobile device and the head-mounted display. Furthermore, the camera in the videos remained in a static position. Despite this, the motion sickness on the head-mounted display was still rated significantly higher, further substantiating H_3.

5.1.2 Methods

The performance of the three methods used in the pre-study did confirm H_4. The gestures featured in this study were used under ideal circumstances: The video was static with only the woman standing in the room moving and also facing the camera. The gestures were repeated five times and after a few video, the participants in the study knew what to look for as the questions were the same after each video. Still, the differences were significant and facial expressions were outperformed by the other gestures in all performance categories and on all devices. Small gestures performed better albeit worse than the big gestures. This was all to be expected, however, the differences under these conditions show that facial expressions are not a practical way to guide users. If the person performing the expressions does not face the camera and the distance to the camera is bigger, the user will most likely be unable to follow the guidance even if it is perceived as such. Small and big gestures were deemed to have the potential to guide users under the circumstances so that they were examined further in the main study.

5.2 Main Study

The main study was designed to gather information on a wide range of visual guidance methods on two types of devices.

5.2.1 Devices

The differences in performance on the two devices were not significant. Still, more users were able to identify the indicated object and the method of guidance when using then head-mounted display. This shows a slight tendency to confirm H_3. In terms of preference, the overall task load was on the same level while the overall user experience was significantly better on the head-mounted display. This is also true for the immersion rating. However, motion sickness was a smaller problem on the mobile device. Overall, the head-mounted display provided the better experience but users had to handle the motion sickness. Additionally, immersion was one of the key motivations for this study. In combination with the novelty rating, which was far better on the head-mounted display, and a less confusing experience, the head-mounted display can be rated superior to the mobile device and was also preferred by the users.
None of the participants suffered from symptoms induced by motion sickness that made them stop watching a video although at least one participant stated to suffer from motion sickness even when using a regular computer to play certain games. A camera that is not static might multiply the effects seen in this study.
A note on the significant difference in the performance rating of the task load index: The standard design of the questionnaire that was also used in this study rates most categories on a scale from "very low" to "very high". The performance rating is the only exception as it is rated from "perfect" to "failure". This caused confusion and some participants stated to have made a mistake in this rating. This was corrected when possible, however, the results may not be reliable.

5.2.2 Methods

H_0 states that every method of guidance improves the performance. As all methods tested in this study perform better than the videos with no guidance, this hypothesis is true.

No Guidance

Using no guidance makes it very difficult for the user to notice an area of interest that is not in the field of view at a given point. In the study, 12.5 % of the user stated that something was indicated. Half of them saw the manipulated object by chance when they were looking around in the virtual environment. The others misinterpreted something, e.g. a normal gesture in the video. This shows that guidance is necessary if something is supposed to be seen by the user and not just by chance.

Interestingly, the user experience was rated quite low with only forced rotation performing worse. This was not expected, instead, a lower user experience was expected from the addition of methods. However, this could be because the participants knew that something would be indicated. As no guidance occurred and no object was spotted by the participant, the user experience was rated low.
Despite not being a significant difference, the immersion rating for no guidance was the lowest of all methods. This is surprising as all methods add some kind of interaction with the user for guidance which should be more prone to a break in immersion than no guidance.

Forced Rotation

Forced rotation can be implemented when the video itself is already finished. This is most likely the reason for Facebook's implementation as it allows users to add guidance to arbitrary 360-degree videos. However, this leads to a number of disadvantages.
First of all, this implementation is unique to Facebook's player and cannot be transferred as it is not part of the video. Adding the rotation during post-production, like it was done in the study, will lead to inaccurate rotations as the user's exact field of view is not known.
In addition to these problems, the method itself suffers from several flaws. Some participants interpreted the rotation in the video as an error. One stated that the tracking of the head-mounted display must have been erroneous and others did not identify the rotation as a method at all. A common behaviour was the rotation of the video by the participants to compensate for the rotation from the method so that their perceived area of interest (the game being played) remained in their gaze. This confirms similar observations by Danieau, Guillo, and Doré, 2017 in their implementation of the method. These factors explain why less than 40 % were able to identify the manipulated object despite forced rotation being the only method that does not require the user to actively look to follow the guidance.
The rotation does also diminish the user's freedom in the virtual environment, represented in a low immersion rating in the study. This is a probable factor in Facebook's decision to make their implementation optional. The possibility to activate or deactivate the guidance is an advantage of the method.
Forced rotation offers the worst user experience of the tested methods and did also induce the most motion sickness. This was expected. It was the only method which led users to state that they felt unpleasant.
A relatively easy implementation and a decent performance in comparison to the other tested methods are the advantages of forced rotation. However, the downsides are immense: Reduced freedom and low immersion, a lousy user experience, a high probability of at least some motion sickness, the confusion of the user and a clear intrusion into the viewing experience are all obvious problems of this method.

Object to Follow

The vast majority of users recognised the sphere in the videos as an indication. However, more than a third of these users did not follow the guidance. Three main reason could be observed in the study: Some users identified the guidance but decided not to follow it because the narrative was deemed more important. Others tried to follow the sphere but were unable to do so because of the sphere's velocity or imprecise physical movement. The third group did notice the sphere, but did not notice the guidance so that their gaze stayed on the narrative. While it contributes to the user's freedom when the guidance is optional, the other two behaviours mentioned could be flaws in the implementation and not the method.
Despite these problems, only one method was more successful in guiding the users to the manipulated object (person to follow). The task load was rated lowest and the user experience highest of all methods with no impact on motion sickness and no significant impact on immersion.
Object to follow has the potential to guide users in an unobtrusive way, but the implementation is crucial.

Person to Follow

Almost half of the participants that were able to name the manipulated object did not identify the person in the video as the method of guidance. Additionally, the immersion was rated best. This leads to the conclusion that following a person through the scene is a very subtle and natural way of guiding the user. Furthermore, nearly two thirds of the users were able to name the manipulated object, making the method the most successful in this category. The best user experience in the study and low ratings in motion sickness and task load show a clear preference for this method.
Although the method performed very well in this study, implementation into a scene could be very complex. A person acting as the guide is required in the scene, that person has to be free to move around while maintaining the focus of the user and ideally, the movement is motivated by the narrative of the scene. While the implementation in hide-and-seek in the main study was a good example how this can be done, implementation would have been far more difficult in all other games.
Person to follow performed best in this study, however, implementing the method in a scene with a already existing narrative can be very complicated.

Object Manipulation

Apart from no guidance, object to follow performed worst in the study. The main reason for this is the limited area in which the manipulation is visible. The method aims to increase the size of the area of interest to gain the attention of the user and is successful in doing so as the comparison to no guidance proves. However, even these enlarged manipulations remain fruitless when the user's gaze is oriented in a completely different direction.

All this was to be expected and the method did perform decent in the questionnaires which was not surprising as well because the method remained invisible to more than three quarters of the user. This means that object manipulation could be an easy way to increase the size of the area of interest, albeit only to a certain degree. This could be useful when the expected position of the user's gaze is already close to the area of interest. Observations from the main study support this behaviour.

Environment Manipulation

Despite altering the appearance of almost the whole scene, less than two thirds of the participants recognised this as guidance. Similar to forced rotation, some users just accepted the manipulation and kept watching without any kind of reorientation. Others regarded the reduced brightness as a notice from the smartphone and tapped the display to prevent it from turning off. These numbers and behaviours clearly show that the method caused confusion. This is also evident as forced rotation was the only method rated more confusing.
As the method offers several ways to manipulate the image in varying degrees of intensity, a different implementation of the method could alter the performance significantly. This is also true for the animations used to introduce the visual effects of the method.
Apart from this, the vast manipulations of the scene made the videos slightly more prone to motion sickness and immersion was rated below most other methods.
Overall, environment manipulation did not perform well in the study. This could be improved by using a different implementation. However, Danieau, Guillo, and Doré, 2017 did use similar manipulations to guide users and came to the conclusion that "[i]mplicitly driving the user's gaze is an ambitious challenge and does not seem really possible within a duration of few seconds" (Danieau, Guillo, and Doré, 2017, p. 206).

Gestures

During the study, less than half of the participants were able to name the gestures as a method of guidance. This was true for both types of gestures. Despite the same number of participants noticing the indication, the small gestures were slightly more successful in guiding the users to the manipulated object. Considering the results from the pre-study, this was not expected. An explanation might be that the gestures were performed under less ideal conditions compared to the first study. The general direction indicated by the small gestures might have been sufficient to guide users to the area of interest. Additionally, the indicated object was manipulated whereas the objects in the pre-study had to be identified by the participants without further manipulations.
This could mean that in practice, the small gestures are good enough for guidance if a rough direction is indicated but bigger gestures (i.e. pointing) are required for more specific guidance. However, the gestures were not interpreted as guidance by a lot of users. Several participants stated that they did not comprehend that they were meant to follow the gesture and just accepted them as a part of the narrative.

As expected from these methods, no impact on immersion and motion sickness was observed. However, big gestures were rated to be most conventional. This could indicate that the small gestures were perceived as a more subtle approach.
In conclusion, the gestures performed underwhelming. On the one hand this might be because gestures were not clear enough to be considered as guidance. On the other hand less subtle gestures may have an impact on immersion.

5.2.3 Additional Results

H_1 states that methods with less differences to the original video provide a better immersion. As no guidance, person to follow and the gestures do not manipulate the original video at all, the immersion rating should be best in the videos using these methods.
The immersion rating did not yield significant results. However, forced rotation and environment manipulation rated worse than the other methods. As major manipulation are present in the videos using these methods, this supports the hypothesis. Despite this, no guidance was rated even lower in terms of immersion although no method is present.
H_1 cannot be confirmed but an impact on immersion because of the manipulations used by the methods is probable.

The overall user experience was rated very low with no guidance as only forced rotation was rated worse. Due to the absence of any method, this was not expected but could be explained when comparing the rating to the accuracy rating (see Figure 4.38). With the exception of forced rotation, the user experience was rated higher when the accuracy rating and therefore the performance of the participant was better. This indicates an impact on the user experience if the guidance fails to bring the area of interest to the user's attention. These results suggest that H_2 is true.

The study shows that the experience of the user has an impact on their performance. Users who were experience in virtual environments (including virtual reality) performed better than those with no or only limited experience.

The implementation of the method object to follow by L. T. Nielsen et al., 2016 yielded similar results in preference when compared to the main study of this work, but differed in terms of performance. The methods investigated in the study (no guidance, forced rotation and a firefly as an object to follow) all showed a very similar performance. This difference to the results of the main study may be for two reasons in the design of the experiment: Lack of manipulations outside of the narrative and very long durations of the indicated elements. Firstly, the scene consisted of only one narrative whereas the manipulated objects in the main study of this work are elements outside of the narrative. If the objects were part of the story, user would not need guidance. Secondly, the indicated elements of the scene are the only visible action that takes places for a significant amount of time (e.g. a man is standing outside the window for almost one minute).

These differences show that the implementation of the methods and the scene itself are very important for the performance of the method. As L. T. Nielsen et al., 2016 suggest, the main study in this work uses more complex scenes as the manipulated object introduces an additional layer of complexity apart from the main narrative.

6 Conclusion

It is necessary to rethink the techniques and methods established in traditional film when working with 360-degree videos. While the new freedom of the user inside the virtual environments is one of the benefits, it is also creating problems. The methods investigated in this work aim to give the user the guidance that is needed without taking away the freedom.

Subtle guidance is important to maintain the immersion. However, visual guidance has to be perceivable to be able to follow it. The balance between these two factors is difficult to find and not solely answered by the choice of a method. The implementation is crucial as is the scene itself.

The studies show that there is not a single method capable of fulfilling all the requirements. Additionally, most of the methods were unable to guide more than half of the participants to the area of interest. When creating 360-degree videos, it is vital to plan the guidance during pre-production to be able to choose the best implementation. Additionally, the combination of visual guidance with auditory clues could considerably improve guidance. Methods like object manipulation may be more successful when used together with sounds (e.g. a TV set illuminating the room and making noise).

Results show that methods of visual guidance have potential but need to be worked on to keep the aforementioned balance between guidance and immersion. It is therefore interesting to observe how Facebook implemented forced orientation into its system. As a provider of at least hundreds of thousands of 360-degree videos, decisions like this could define the user's reaction to this kind of content as well as establish conventions. It is therefore important to provide alternatives to this kind of guidance as it seems to be inferior to other methods.

Although concrete numbers are unavailable, mobile devices seem to be the most used device category for 360-degree videos. The studies suggest that these devices are not ideal for virtual environments. However, it is crucial to provide users with the best experience possible on all devices, underlining the need for clear guidance.

Guiding users through virtual environments is a difficult task to accomplish. Subtle visual guidance to lead the attention of the user with minimal interference in the narrative of the video is even more challenging. The overview and the studies of this work lay the groundwork for successful implementations and the web application (see Appendix A) demonstrates how this can be used in practice. However, additional work is required to refine the implementations of the methods.

7 Future Work

As mentioned before, user guidance in virtual environments needs additional work to further improve the viewing experience. There is a number of ways for this thesis to serve as a basis.

First of all, concentrating on visual guidance was a necessary step to be able to cover a wide variety of methods. However, sound is an integral part of the viewing experience with great potential to guide users. Future work should combine visual and auditory clues to be able to find better ways to guide users.

Facebook's implementation of forced rotation shows the need for methods that can easily be added to videos. Therefore, refinement of this method is necessary as well as work on other methods which can be added after the video itself is already finished.

The object to follow has been refined in several implementations. However, there may be different ways to use it. In Peck, Fuchs, and Whitton, 2009, such an object is used to distract the user. It may be possible to attract attention by distracting the user from the narrative. Further combinations with person to follow are also feasible.

Environment manipulation could also benefit from an improved implementation. Danieau, Guillo, and Doré, 2017 already showed that it is not an easy task to accomplish, but a successful implementation does have the potential to challenge forced rotation as a method to be added after post-production. This could be very valuable despite the method's underwhelming performance in this work's main study.

Gestures as methods tend to be perceived as part of the video instead of guidance. If this can be changed and users intuitively react to the gestures, this becomes a valuable method.

All of the implementations of the methods used in this work are examples of how the methods could work. Every method on its own has the potential of a variety of implementations which could drastically change both viewing experience and guiding capabilities. It is crucial to find the best ways to use a method in order to be able to finally judge its eligibility as a method of user guidance.

A more immersive experience does seem to be also more prone to motion sickness. Further improvement of the hardware used for virtual reality is therefore necessary. This does include the addition of new techniques such as eye-tracking which also creates new possibilities for guidance.

In conclusion, future work is necessary to improve methods of guidance and the hardware used to view virtual environments. This work offers an outline to build upon.

8 Further Reading

Additional material related to this work is diverse. The Filmmaker's Handbook (Ascher and Pincus, 2012) and Film Art: An Introduction (Bordwell and Thompson, 2008) introduce techniques used in traditional film. While these may not always be compatible with virtual environments (see section 2.2), the established standards of film making do also apply to narratives in virtual environments and expecially 360-degree videos. A look back on the history of spherical videos and their technical details is presented in Surround video: a multihead camera approach (F. Nielsen, 2005).
Hamlet on the Holodeck: The Future of Narrative in Cyberspace (Murray, 1997) details the changes in storytelling when adapted to modern technology. Project Orpheus A Research Study into 360° Cinematic VR (Vosmeer and Schouten, 2017) explores how a traditional television show may be combined with an immersive experience in virtual reality.
Attention: Theory and Practice (Johnson and Proctor, 2004) is a comprehensive look at attention in general and can provide valuable insight to create or refine methods of user guidance.
Movement in virtual realities does induce problems with the available physical space and reorientation techniques or other ways of movement are necessary. Point and Teleport Locomotion Technique Point and Teleport Locomotion Technique for Virtual Reality (Bozgeyikli et al., 2016) and Evaluation of Reorientation Techniques and Distractors for Walking in Large Virtual Environments (Peck, Fuchs, and Whitton, 2009) cover these problems. The approaches that are used are also relevant to 360-degree videos, as shown by the sphere introduced by Peck, Fuchs, and Whitton, 2009 which was used to conceive the firefly in Missing the Point: An Exploration of How to Guide Users' Attention During Cinematic Virtual Reality (L. T. Nielsen et al., 2016).
3D User Interfaces: Theory and Practice (Bowman et al., 2004) is about three-dimensional user interfaces that could contribute to the development of new methods to guide users in virtual reality.
User Experience of Panoramic Video in CAVE-like and Head Mounted Display Viewing Conditions (Philpot et al., 2017) does compare head-mounted displays to CAVE-like systems (Cave Automatic Virtual Environment) where users do not wear a device and are instead seated in a room whose walls serve as a canvas.
Lee, Han, and Choi, 2016 present an algorithm to match the movement of chairs in a 4D cinema to the action of the screen. Adding physical movement to 360-degree videos could provide new methods of user guidance.

A Web Application

To give a practical example of the classification system introduced in Methods of User Guidance, chapter 3, a small single-page web application has been created. This application allows users to weight the categories introduced in this work in order to to find a suitable method of guidance.

The methods were rated on a scale from 0 to 4 in each category. This was done before the user studies and the data set was refined once the studies were completed. It is possible to switch between these data sets. It is also possible to exclude methods that require a person in the scene.

Of course this rating is just a rough approximation and the implementation of the methods remains crucial. Still, this application demonstrates how the complex subject of user guidance in 360-degree videos may be simplified in practice. It could help content creators who are not accustomed with virtual environments to use guidance and improve the result.

The web application is attached to the physical copy of this method and is also available online at https://tiny.cc/OverThereWebApp.

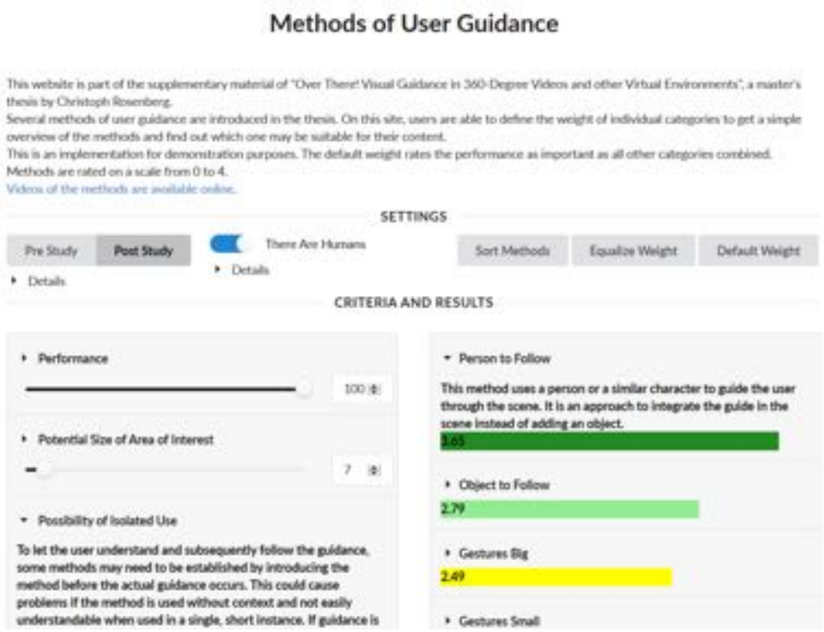

FIGURE A.1: Web application

B Study Questionnaires

B.1 Pre-Study Questionnaires

Participants were instructed to notify the experimenter as soon as they knew what was indicated to them. They were asked to state the object that was indicated and the unique detail of the object. The search time and the number of gestures in the video were recorded.
After all videos had been completed, participants were asked to fill out a number of questionnaires. As the pre-study was conducted in German, the questionnaires are in German as well.

B.1.1 NASA Task Load Index

The NASA Task Load Index was not changed from the standard implementation for the pre-study. See So, 2017 for details.

Klicken Sie in jeder Skale auf den Punkt, der Ihre Erfahrung im Hinblick auf die Aufgabe am besten verdeutlicht. (scales from 0 to 100; increment: 5)
Geistige Anforderung: Gering - Hoch
Körperliche Anforderung: Gering - Hoch
Zeitliche Anforderung: Gering - Hoch
Leistung: Gut - Schlecht
Anstrengung: Gering - Hoch
Frustration: Gering - Hoch

Klicken Sie auf die Dimension, die den jeweils wichtigeren Beitrag zur Arbeitsbelastung hinsichtlich der Aufgabe darstellt. (Users were asked to choose one out of two categories from above; this question was reapeated.)

B.1.2 User Experience Questionnaire

The standard user experience questionnaire from Laugwitz, Held, and Schrepp, 2008 was used with Likert scales from 1 to 7.

Um Ihre Erfahrung zu bewerten, füllen Sie bitte den nachfolgenden Fragebogen aus. Er besteht aus Gegensatzpaaren von Eigenschaften, die ihre Erfahrung haben kann. Abstufungen zwischen den Gegensätzen sind durch Kreise dargestellt. Durch Ankreuzen eines dieser Kreise können Sie Ihre Zustimmung zu einem Begriff äußern.
Entscheiden Sie möglichst spontan. Es ist wichtig, dass Sie nicht lange über die Begriffe nachdenken, damit Ihre unmittelbare Einschätzung zum Tragen kommt.
Bitte kreuzen Sie immer eine Antwort an, auch wenn Sie bei der Einschätzung zu einem Begriffspaar unsicher sind oder finden, dass es nicht so gut passt.
Es gibt keine „richtige“ oder „falsche“ Antwort. Ihre persönliche Meinung zählt!
unerfreulich - erfreulich
unverständlich - verständlich
kreativ - phantasielos
leicht zu lernen - schwer zu lernen
wertvoll - minderwertig
langweilig - spannend
uninteressant - interessant
unberechenbar - voraussagbar
schnell - langsam
originell - konventionell
behindernd - unterstützend
gut - schlecht
kompliziert - einfach
abstoßend - anziehend
herkömmlich - neuartig
unangenehm - angenehm
sicher - unsicher
aktivierend - einschläfernd
erwartungskonform - nicht erwartungskonform
ineffizient - effizient
übersichtlich - verwirrend
unpragmatisch - pragmatisch
aufgeräumt - überladen
attraktiv - unattraktiv
sympathisch - unsympathisch
konservativ - innovativ

B.1.3 Motion Sickness Awareness Questionnaire

The Motion Sickness Awareness Questionnaire from J. Gianaros et al., 2001 was used. (Likert scales from 1 to 9)

Mir war schlecht
Ich fühlte mich kurz vor einer Ohnmacht
Ich fühlte mich genervt/irritiert
Ich fühlte mich verschwitzt
Ich fühlte mich mulmig/übel
Ich fühlte mich benommen
Ich fühlte mich schläfrig
Ich fühlte mich klamm/kalten Schweiß
Ich fühlte mich desorientiert
Ich fühlte mich müde/ermüdet
Ich fühlte mich angeekelt/angewidert
Mir wurde warm/heiß
Mir wurde schwindlig
Ich fühlte mich, als würde sich alles drehen
Ich fühlte mich, als müsste ich mich übergeben
Ich fühlte mich unbehaglich

B.1.4 Immersion Questionnaire

The immersion questionnaire from Slater, Usoh, and Steed, 1994 was used. (Likert scales from 1 to 7 and a comment area)

- Bitte bewerten Sie, inwiefern Sie sich gefühlt haben, als wären Sie in dem Raum gewesen, anhand der der folgenden Skala von 1 bis 7, wobei 7 für Ihr normales Gefühl sich in einem Raum zu befinden steht.
- In welchem Maße gab es Zeiten, in denen der Raum für Sie die Realität war?
- Wenn Sie an Ihre Erfahrung zurückdenken, denken Sie an den Raum eher als Bilder, die Sie gesehen haben oder eher als Ort, den Sie besucht haben?
- Welches Gefühl hat für Sie während des Experiments insgesamt überwogen: das Gefühl in dem Raum zu sein oder das Gefühl woanders zu sein?

Betrachten Sie Ihre Erinnerungen an den Raum. Wie ähnlich in Bezug auf die Gedächtnisstruktur ist sie im Vergleich zur Gedächtnisstruktur, die Sie von anderen Orten haben, die Sie heute besucht haben? Zur Gedächtnisstruktur gehören Dinge wie das Ausmaß der bildlichen Erinnerung an den Raum, ob Ihre Erinnerung in Farbe ist, inwiefern die Erinnerung lebendig oder realistisch ist, ihre Größe, Ort in Ihrer Vorstellung, inwiefern sie panoramaartig in Ihrer Vorstellung ist und weitere solcher strukturellen Elemente.
- Haben Sie während des Experiments oft gedacht, Sie wären tatsächlich in dem Raum?
- Bitte hinterlassen Sie weitere Anmerkungen über Ihre Erfahrung. Insbesondere: Welche

Dinge haben Ihnen geholfen, sich so zu fühlen, als wären Sie tatsächlich in dem Raum und welche Dinge haben Sie aus diesem Gefühl herausgerissen?

B.1.5 Demographic Questionnaire

- Sind Sie farbenblind? Ja - Nein
- Benötigen Sie eine Sehhilfe? Ja - Nein
- Sind Sie in Ihrer Bewegung eingeschränkt? Ja - Nein
- Ich habe bereits positive Erfahrungen mit 360°-Videos auf folgenden Geräten gemacht: [Desktop PC/Notebook] [Smartphone/Tablet] [VR-Brille] (Lickert scales)
- Welches Ausgabegerät würden Sie bei den folgenden Aktivitäten bevorzugen? [Film schauen] [Spiel spielen] [Buch/Zeitung lesen] [Arbeiten] [Lernen] (Participants could choose from: Smartphone, Tablet, - - Desktop PC/Notebook, TV, VR-Brille)
- 360°-Videos: Welche Erfahrungen haben Sie und wie viele?
- Haben Sie bereits Erfahrung mit Virtual Reality, z.B. aus Spielen oder anderen Studien? Welche?
- Alter
- Geschlecht

B.2 Main Study Questionnaires

The main study was conducted in German and English. Participants could choose their preferred language. Here, only the English questions are listed for the questionnaires as those are also part of the pre-study (see section B.1).

B.2.1 Questions After a Video

After every video in the main study, participants were asked the following question:
Has something been indicated to you?/Wurden Sie auf etwas hingewiesen?
If this question was answered positively, participants were asked: What has been indicated?/Worauf wurden Sie hingewiesen?
How was this indicated to you?/Wie wurden Sie darauf hingewiesen?

Additionally, after every video, participants answered these questions based on the questionnaire by So, 2017, Laugwitz, Held, and Schrepp, 2008, J. Gianaros et al., 2001 and Slater, Usoh, and Steed, 1994.

Please rate your experience.
You are rating your experience with the last video.
Please decide spontaneously. Don't think too long about your decision to make sure that you convey your original impression.
It is your personal opinion that counts. Please remember: there is no wrong or right answer! (Likert scales from 1 to 7)
unpleasant - pleasant
inefficient - efficient
clear - confusing
obstructive - supportive
valuable - inferior
conventional - inventive

(Likert scales from 1 to 9)
I felt sick to my stomach: Not at all - Severely
I felt sweaty: Not at all - Severely
I felt disoriented or dizzy: Not at all - Severely
I felt tired/fatigued: Not at all - Severely

(Likert scale from 1 to 7)
I had a sense of "being there" in the room. Please rate your sense of being in the shopping environment, on the following scale from 1 to 7, where 7 represents your normal experience of being in a place. Not at all - Very much
Please write down any further comments that you wish to make about your experience. In particular, what things helped to give you a sense of 'really being' in the room, and what things acted to 'pull you out' of this?

(Likert scales from 0 to 10)
Mental Demand: Low - High
Physical Demand: Low - High
Temporal Demand: Low - High
Performance: Good - Poor
Effort: Low - High
Frustration: Low - High

B.2.2 Questions After the Experiment

After all videos were completed, participants were asked to fill out the weighting part of the NASA Task Load Index and a demographic questionnaire.

Click on the factor that represents the more important contributor to workload for the task. (Two factors from the Task Load Index were display; this question was repeated)
- How often do you watch 360-degree videos? Never - Very often (Likert scale from 1 to 5)
- 360-degree videos: How experienced are you and why?
- Virtual reality: How experienced are you and why?
- My experience with 360-degree videos on the following devices was positive: [Desktop PC/Notebook] [Smartphone/Tablet] [VR Headset] Strongly disagree - Strongly agree, No experience (Likert scale from 1 to 5)
- Do you prefer 360-degree videos to traditional videos? - In which areas will 360-degree videos replace traditional videos? - I find it annoying if my attention is guided in videos.
- Are you color-blind? - Do you need optical aids (glasses, contact lenses)? - Is your motion impaired? - Age - Sex

Bibliography

Ascher, Steven and Edward Pincus (2012). *The Filmmaker's Handbook: A Comprehensive Guide for the Digital Age: 2013 Edition*. Penguin Publishing Group. ISBN: 9781101613801.

Bailey, Reynold et al. (2009). "Subtle Gaze Direction". In: *ACM Trans. Graph.* 28.4, 100:1–100:14. ISSN: 0730-0301. DOI: 10.1145/1559755.1559757. URL: http://doi.acm.org/10.1145/1559755.1559757.

Beddoe-Stephens, Paul (2016a). *Facebook 360 Updates and introducing Surround 360*. URL: https://media.fb.com/2016/04/12/facebook-360-updates/ (visited on Sept. 19, 2017).

— (2016b). *New Publisher Tools for 360 Video*. URL: https://media.fb.com/2016/08/10/new-publisher-tools-for-360-video/ (visited on Sept. 19, 2017).

Bordwell, D. and K. Thompson (2008). *Film art: An Introduction*. McGraw Hill. ISBN: 9780073310275.

Bowman, D. et al. (2004). *3D User Interfaces: Theory and Practice*. Pearson Education. ISBN: 9780133390568.

Bozgeyikli, Evren et al. (2016). "Point & Teleport Locomotion Technique for Virtual Reality". In: *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*. CHI PLAY '16. Austin, Texas, USA: ACM, pp. 205–216. ISBN: 978-1-4503-4456-2. DOI: 10.1145/2967934.2968105. URL: http://doi.acm.org/10.1145/2967934.2968105.

Brooks, Frederick P. (1999). "What's Real About Virtual Reality?" In: *IEEE Comput. Graph. Appl.* 19.6, pp. 16–27. ISSN: 0272-1716. DOI: 10.1109/38.799723. URL: http://dx.doi.org/10.1109/38.799723.

Chen, Shenchang Eric (1995). "QuickTime VR: An Image-based Approach to Virtual Environment Navigation". In: *Proceedings of the 22Nd Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH '95. New York, NY, USA: ACM, pp. 29–38. ISBN: 0-89791-701-4. DOI: 10.1145/218380.218395. URL: http://doi.acm.org/10.1145/218380.218395.

Chittaro, Luca and Stefano Burigat (2004). "3D Location-pointing As a Navigation Aid in Virtual Environments". In: *Proceedings of the Working Conference on Advanced Visual Interfaces*. AVI '04. Gallipoli, Italy: ACM, pp. 267–274. ISBN: 1-58113-867-9. DOI: 10.1145/989863.989910. URL: http://doi.acm.org/10.1145/989863.989910.

Danieau, F., A. Guillo, and R. Doré (2017). "Attention guidance for immersive video content in head-mounted displays". In: *2017 IEEE Virtual Reality (VR)*, pp. 205–206. DOI: 10.1109/VR.2017.7892248.

Davis, Simon, Keith Nesbitt, and Eugene Nalivaiko (2014). "A Systematic Review of Cybersickness". In: *Proceedings of the 2014 Conference on Interactive Entertainment*. IE2014. Newcastle, NSW, Australia: ACM, 8:1–8:9. ISBN: 978-1-4503-2790-9. DOI: 10.1145/2677758.2677780. URL: http://doi.acm.org/10.1145/2677758.2677780.

Google (2016). *Spatial Audio*. URL: https://developers.google.com/vr/concepts/spatial-audio (visited on Sept. 19, 2017).

— (2017). *Tabel - Making Of*. URL: https://tabel.withgoogle.com/making-of (visited on Sept. 19, 2017).

Hata, Hajime, Hideki Koike, and Yoichi Sato (2016). "Visual Guidance with Unnoticed Blur Effect". In: *Proceedings of the International Working Conference on Advanced Visual Interfaces*. AVI '16. Bari, Italy: ACM, pp. 28–35. ISBN: 978-1-4503-4131-8. DOI: 10.1145/2909132.2909254. URL: http://doi.acm.org/10.1145/2909132.2909254.

Hong, Seokjun and Gerard J. Kim (2016). "Accelerated Viewpoint Panning with Rotational Gain in 360 Degree Videos". In: *Proceedings of the 22Nd ACM Conference on Virtual Reality Software and Technology*. VRST '16. Munich, Germany: ACM, pp. 303–304. ISBN: 978-1-4503-4491-3. DOI: 10.1145/2993369.2996309. URL: http://doi.acm.org/10.1145/2993369.2996309.

Hosseini, Mohammad and Viswanathan Swaminathan (2016). "Adaptive 360 VR Video Streaming: Divide and Conquer!" In: *CoRR* abs/1609.08729. URL: http://arxiv.org/abs/1609.08729.

J. Gianaros, Peter et al. (2001). "A Questionnaire for the Assessment of the Multiple Dimensions of Motion Sickness". In: 72, pp. 115–9.

Johnson, Addie and Robert W. Proctor (2004). *Attention: Theory and Practice*. SAGE Publications, Inc. ISBN: 9780761927600.

Korolov, Maria (2016). *Report: 98% of VR headsets sold this year are for mobile phones*. URL: http://www.hypergridbusiness.com/2016/11/report-98-of-vr-headsets-sold-this-year-are-for-mobile-phones/ (visited on Sept. 19, 2017).

Laugwitz, Bettina, Theo Held, and Martin Schrepp (2008). "Construction and Evaluation of a User Experience Questionnaire". In: *Proceedings of the 4th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society on HCI and Usability for Education and Work*. USAB '08. Graz, Austria: Springer-Verlag, pp. 63–76. ISBN: 978-3-540-89349-3. DOI: 10.1007/978-3-540-89350-9_6. URL: http://dx.doi.org/10.1007/978-3-540-89350-9_6.

Lee, Jaebong, Bohyung Han, and Seungmoon Choi (2016). "Interactive Motion Effects Design for a Moving Object in 4D Films". In: *Proceedings of the 22Nd ACM Conference on Virtual Reality Software and Technology*. VRST '16. Munich, Germany: ACM, pp. 219–228. ISBN: 978-1-4503-4491-3. DOI: 10.1145/2993369.2993389. URL: http://doi.acm.org/10.1145/2993369.2993389.

Mackworth, Norman H. and Anthony J. Morandi (1967). "The gaze selects informative details within pictures". In: *Perception & Psychophysics* 2.11, pp. 547–552. ISSN: 1532-5962. DOI: 10.3758/BF03210264. URL: https://doi.org/10.3758/BF03210264.

Mateescu, Victor A. and Ivan V. Bajić (2014). "Attention Retargeting by Color Manipulation in Images". In: *Proceedings of the 1st International Workshop on Perception Inspired Video Processing*. PIVP '14. Orlando, Florida, USA: ACM, pp. 15–20. ISBN: 978-1-4503-3125-8. DOI: 10.1145/2662996.2663009. URL: http://doi.acm.org/10.1145/2662996.2663009.

McNamara, Ann, Reynold Bailey, and Cindy Grimm (2008). "Improving Search Task Performance Using Subtle Gaze Direction". In: *Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization*. APGV '08. Los Angeles, California:

ACM, pp. 51–56. ISBN: 978-1-59593-981-4. DOI: 10.1145/1394281.1394289. URL: http://doi.acm.org/10.1145/1394281.1394289.

McNamara, Ann, Thomas Booth, et al. (2012). "Directing Gaze in Narrative Art". In: *Proceedings of the ACM Symposium on Applied Perception*. SAP '12. Los Angeles, California: ACM, pp. 63–70. ISBN: 978-1-4503-1431-2. DOI: 10.1145/2338676.2338689. URL: http://doi.acm.org/10.1145/2338676.2338689.

Murray, Janet Horowitz (1997). *Hamlet on the Holodeck: The Future of Narrative in Cyberspace*. New York, NY, USA: The Free Press. ISBN: 0684827239.

Neng, Luís A. R. and Teresa Chambel (2010). "Get Around 360° Hypervideo". In: *Proceedings of the 14th International Academic MindTrek Conference: Envisioning Future Media Environments*. MindTrek '10. Tampere, Finland: ACM, pp. 119–122. ISBN: 978-1-4503-0011-7. DOI: 10.1145/1930488.1930512. URL: http://doi.acm.org/10.1145/1930488.1930512.

Nielsen, Frank (2005). "Surround video: a multihead camera approach". In: *The Visual Computer* 21.1, pp. 92–103. ISSN: 1432-2315. DOI: 10.1007/s00371-004-0273-z. URL: https://doi.org/10.1007/s00371-004-0273-z.

Nielsen, Lasse T. et al. (2016). "Missing the Point: An Exploration of How to Guide Users' Attention During Cinematic Virtual Reality". In: *Proceedings of the 22Nd ACM Conference on Virtual Reality Software and Technology*. VRST '16. Munich, Germany: ACM, pp. 229–232. ISBN: 978-1-4503-4491-3. DOI: 10.1145/2993369.2993405. URL: http://doi.acm.org/10.1145/2993369.2993405.

Peck, T. C., H. Fuchs, and M. C. Whitton (2009). "Evaluation of Reorientation Techniques and Distractors for Walking in Large Virtual Environments". In: *IEEE Transactions on Visualization and Computer Graphics* 15.3, pp. 383–394. ISSN: 1077-2626. DOI: 10.1109/TVCG.2008.191.

Philpot, Adam et al. (2017). "User Experience of Panoramic Video in CAVE-like and Head Mounted Display Viewing Conditions". In: *Proceedings of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video*. TVX '17. Hilversum, The Netherlands: ACM, pp. 65–75. ISBN: 978-1-4503-4529-3. DOI: 10.1145/3077548.3077550. URL: http://doi.acm.org/10.1145/3077548.3077550.

Ramalho, João and Teresa Chambel (2013). "Windy Sight Surfers: Sensing and Awareness of 360° Immersive Videos on the Move". In: *Proceedings of the 11th European Conference on Interactive TV and Video*. EuroITV '13. Como, Italy: ACM, pp. 107–116. ISBN: 978-1-4503-1951-5. DOI: 10.1145/2465958.2465969. URL: http://doi.acm.org/10.1145/2465958.2465969.

Raza Abidi, Syed Shaukat, MaryAnn Williams, and Benjamin Johnston (2013). "Human Pointing As a Robot Directive". In: *Proceedings of the 8th ACM/IEEE International Conference on Human-robot Interaction*. HRI '13. Tokyo, Japan: IEEE Press, pp. 67–68. ISBN: 978-1-4673-3055-8. URL: http://dl.acm.org/citation.cfm?id=2447556.2447570.

Saba, Maher (2015). *Introducing 360 Video on Facebook*. URL: https://newsroom.fb.com/news/2015/09/introducing-360-video-on-facebook/ (visited on Sept. 19, 2017).

Slater, Mel, Martin Usoh, and Anthony Steed (1994). "Depth of Presence in Virtual Environments". In: *Presence: Teleoper. Virtual Environ.* 3.2, pp. 130–144. ISSN: 1054-7460. DOI: 10.1162/pres.1994.3.2.130. URL: http://dx.doi.org/10.1162/pres.1994.3.2.130.

So, Phil (2017). *TLX @ NASA Ames*. URL: https://humansystems.arc.nasa.gov/groups/TLX/index.php (visited on Sept. 19, 2017).

Sridharan, Srinivas and Reynold Bailey (2016). "Saliency and Optical Flow for Gaze Guidance in Videos". In: *Proceedings of the ACM Symposium on Applied Perception*. SAP '16. Anaheim, California: ACM, pp. 140–140. ISBN: 978-1-4503-4383-1. DOI: 10.1145/2931002.2948725. URL: http://doi.acm.org/10.1145/2931002.2948725.

Sridharan, Srinivas, James Pieszala, and Reynold Bailey (2015). "Depth-based Subtle Gaze Guidance in Virtual Reality Environments". In: *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception*. SAP '15. Tübingen, Germany: ACM, pp. 132–132. ISBN: 978-1-4503-3812-7. DOI: 10.1145/2804408.2814187. URL: http://doi.acm.org/10.1145/2804408.2814187.

Suma, Evan A. et al. (2012). "A Taxonomy for Deploying Redirection Techniques in Immersive Virtual Environments". In: *Proceedings of the 2012 IEEE Virtual Reality*. VR '12. Washington, DC, USA: IEEE Computer Society, pp. 43–46. ISBN: 978-1-4673-1247-9. DOI: 10.1109/VR.2012.6180877. URL: https://doi.org/10.1109/VR.2012.6180877.

Van den Broeck, Marc, Fahim Kawsar, and Johannes Schöning (2017). *It's All Around You: Exploring 360° Video Viewing Experiences on Mobile Devices*. 2017 ACM on Multimedia Conference. DOI: 10.1145/3123266.3123347. URL: https://doi.org/10.1145/3123266.3123347.

Verma, Sanjeev (2015). *A new way to see and share your world with 360-degree video*. URL: https://youtube-creators.googleblog.com/2015/03/a-new-way-to-see-and-share-your-world.html (visited on Sept. 19, 2017).

Vosmeer, Mirjam and Ben Schouten (2017). "Project Orpheus A Research Study into 360° Cinematic VR". In: *Proceedings of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video*. TVX '17. Hilversum, The Netherlands: ACM, pp. 85–90. ISBN: 978-1-4503-4529-3. DOI: 10.1145/3077548.3077559. URL: http://doi.acm.org/10.1145/3077548.3077559.

Wonner, Jonathan et al. (2013). "Bubble Bee, an Alternative to Arrow for Pointing out Directions". In: *Proceedings of the 19th ACM Symposium on Virtual Reality Software and Technology*. VRST '13. Singapore: ACM, pp. 97–100. ISBN: 978-1-4503-2379-6. DOI: 10.1145/2503713.2503753. URL: http://doi.acm.org/10.1145/2503713.2503753.

YouTube (2017). *Use spatial audio in 360-degree and VR videos*. URL: https://support.google.com/youtube/answer/6395969 (visited on Sept. 19, 2017).