Where to Look: Exploring Peripheral Cues for Shifting Attention to Spatially Distributed Out-of-View Objects

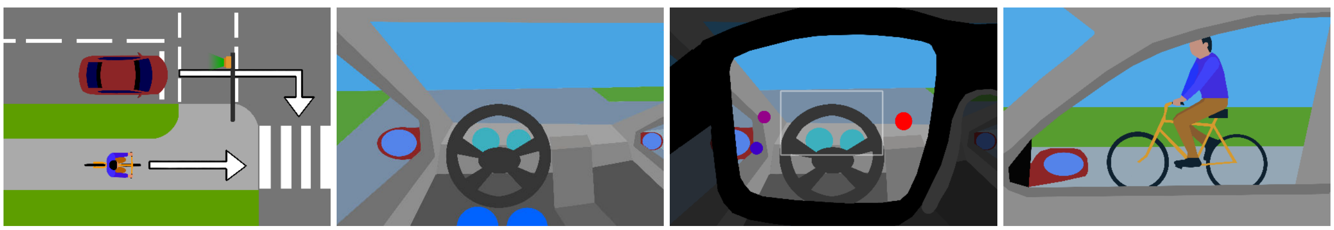
Uwe Gruenefeld
University of Oldenburg
Oldenburg, Germany
uwe.gruenefeld@uol.de

Andreas Löcken
University of Applied Science
Ingolstadt, Germany
andreas.loecken@thi.de

Yvonne Brueck
University of Oldenburg
Oldenburg, Germany
yvonne.brueck@uol.de

Susanne Boll
University of Oldenburg
Oldenburg, Germany
susanne.boll@uol.de

Wilko Heuten
OFFIS - Institute for
Information Technology
Oldenburg, Germany
wilko.heuten@offis.de



(a) Traffic scenario. (b) Driver perspective. (c) Visualization and cue. (d) Shifting target.

Figure 1: **An example to demonstrate our approach of using visual cues to shift attention. *Best seen in color.***

ABSTRACT
Knowing the locations of spatially distributed objects is important in many different scenarios (e.g., driving a car and being aware of other road users). In particular, it is critical for preventing accidents with objects that come too close (e.g., cyclists or pedestrians). In this paper, we explore how peripheral cues can shift a user's attention towards spatially distributed out-of-view objects. We identify a suitable technique for visualization of these out-of-view objects and explore different cue designs to advance this technique to shift the user's attention. In a controlled lab study, we investigate non-animated peripheral cues with audio stimuli and animated peripheral cues without audio stimuli. Further, we looked into how user's identify out-of-view objects. Our results show that shifting the user's attention only takes about 0.86 seconds on average when animated stimuli are used, while shifting the attention with non-animated stimuli takes an average of 1.10 seconds.

AutomotiveUI '18, September 23–25, 2018, Toronto, ON, Canada

ISBN 978-1-4503-5946-7/18/09. . . $15.00

DOI: https://doi.org/10.1145/3239060.3239080

Author Keywords
Attention shift; out-of-view; visualization; peripheral cues; head-mounted; virtual reality.

CCS Concepts
•**Human-centered computing** → *Virtual reality;*

INTRODUCTION
Observing our environment comes with several restrictions. Due to our naturally limited field-of-view (FOV), we are able to observe only parts of our environment at a time. Therefore, objects can disappear out of view, requiring head movement to locate them. However, locating these objects is difficult in many different scenarios because they can be hidden behind other objects or it may simply not be clear where to find them. An example would be a driving scenario in which the driver has to locate road users that might cross their path in order to avoid possible collisions (see Figure 1a). Another example is maritime piloting, in which pilots on a ship bridge have to control and monitor several tugboats assisting in the docking process [21]. Besides in the maritime and automotive domains, locating out-of-view objects is also important in computer games in which players need to know the positions of opponents or teammates.

In previous works, several techniques have been developed to visualize the positions of spatially distributed out-of-view objects (e.g., EyeSee360 [10]). These techniques use representations (so called proxies) of out-of-view objects to display their positions in the user's periphery. By visualizing out-of-view objects in the periphery, the user's focus remains undisturbed, allowing them to focus on the primary task (e.g., driving a car). However, we argue that consistently observing out-of-view objects by following their proxies in a peripheral visualization technique at any given time is too demanding and therefore not possible, especially when driving a car is the primary task of the user. On the other hand, knowing the positions of out-of-view objects that are critical to the user (e.g., because of an intersecting route) would help users to improve their reactions and therefore reduce the number of accidents.

To solve the problem of unperceived critical out-of-view objects we could use two different strategies: (1) we could bring the information about the critical object to the user's attention or (2) bring the user's attention to the critical object. However, bringing the information to the user's attention will probably interfere with the user's primary task (e.g., driving a car) because information is overlaying the user's focus. Bringing the user's attention to the critical out-of-view object by shifting their attention can be time consuming (e.g., if the object is behind the user). In our approach, we suggest a third strategy: (3) we shift the user's attention to proxies of critical out-of-view objects in the user's periphery. Thereby, we do not overlay the user's focus, and we reduce the time required to locate objects. Furthermore, we persistently show all relevant out-of-view objects in the user's periphery and highlight proxies that represent critical objects in their visual presentation. This enables users to be aware of their surroundings in general, not only in dangerous situations (e.g., to persistently monitor traffic in the periphery during automated driving).

In this paper, we want to explore how to visually shift the user's attention to proxies shown in the user's periphery (see Figure 1c). Therefore, we investigate non-animated peripheral cues with audio stimuli and animated peripheral cues without audio stimuli for shifting the user's attention. Besides shifting the user's attention, we were also concerned with identifying out-of-view objects. This is relevant in any situation in which the user is not directly taking action but instead needs to delegate it to somebody else. This is, for example, necessary in the docking scenario [21].

The conducted user study can be seen as a first step of exploring how to visually shift attention towards spatially distributed out-of-view objects in the user's periphery. In our study, we specifically investigate scenarios in which the attention is shifted automatically (exogenous attention shift cp. [26]). To ensure that the visualization stays in the user's periphery, our approach uses a head-mounted device. However, since current Augmented Reality devices suffer from small fields-of-view and we wanted to reduce the influencing factors, we conducted our experiment in Virtual Reality as a controlled lab study.

We propose the following contributions:

1. Visual cue designs for shifting the user's attention to peripheral visualized spatially distributed out-of-view objects

2. An exploration of these cue designs in a controlled lab study with head-mounted Virtual Reality.

RELATED WORK

Our related work section builds upon three pillars: (1) previous work on visualization of out-of-view objects, (2) visual/audio cue designs for shifting attention, and (3) cross-modal effects.

Visualization of out-of-view objects

Similar to the problem of objects receding from view is the visualization of off-screen objects on small screen devices (e.g., smartphones). Contextual Views are especially relevant because they overlay only the screen borders with context information [4, 11, 6]. Applied to head-mounted devices this gives the advantage that the user's focus remains undisturbed. This is especially relevant when augmenting the field-of-view of car drivers. One of the first Contextual views was presented by Zellweger et al. [27], who provided contextual information along the borders. However, in this instance users found it difficult to guess the actual positions of the off-screen objects. Therefore, Halo was suggested as an improvement [1]. It uses circles drawn with their centers around the off-screen objects and cuts the border of the screen slightly. However, a problem for Halo is cluttering, which is the accumulation of many Halos in corners. To solve the problem of cluttering Gustafson et al. suggested Wedge [11]. Besides Halo and Wedge, the smaller shape of arrows is used to point towards off-screen objects. Several studies compared Halo with Arrow approaches [4, 13], where Arrows with fixed sizes performed worse than Halo while scaled arrows performed slightly better.

In recent work [9], Gruenefeld et al. adapted Arrow, Halo, and Wedge to head-mounted Augmented Reality. Their results showed that all of these techniques are applicable to head-mounted devices. However, the showed techniques were only able to visualize out-of-view objects with up to 90 degrees in front of the user. Thereafter, Gruenefeld et al. developed a new visualization technique called EyeSee360 [10]. The technique is inspired by EdgeRadar [12], which utilizes the borders of the screen to visualize the positions of off-screen objects. Instead of using screen borders, EyeSee360 uses the periphery of the user and is able to visualize multiple out-of-view objects spatially distributed in 360° around the user. The technique is explained in more detail in section 3: EyeSee360.

Visual cues for attention shift

Of all things that your eyes see at any instant, you are conscious of only those few to which you direct your attention (cp. [15]). But changing where users direct their visual attention is possible by shifting it with visual cues. With regard to Posner and Petersen [22] there are three phases necessary for shifting the user's attention: (1) disengage the current target, (2) shift attention between stimuli, and (3) engage new target.

In the work Experimental Evaluation of an Augmented Reality Visualization for Directing a Car Driver's Attention, Tönnis et al. investigated the use of Heads-Up display at a fixed position for shifting the car driver's attention [25]. They differed between targets in the driver's frame of reference and targets in an exocentric frame of reference. The problem with shifting the user's attention directly towards the target is that objects behind the user need 180 degree head turning. We argue that this is taking too much time and that it can cause serious injuries if the head is in such an orientation during a collision.

Lin et al. [18] investigated guiding gaze in 360° videos on smartphones. They presented two approaches for guiding attention in 360° videos: Auto Pilot (bringing target to viewers) and Visual Guidance (indicating direction of target). They showed that if increased head movement is necessary (e.g., following a sports video), users preferred Auto Pilot. Furthermore, users found it frustrating to shift to a target that was already gone or a part of a scene that already took place (e.g., a tackling in soccer). This highlights the need for accurate visualization of out-of-view objects.

Two approaches utilizing the peripheral vision are: [23] investigating in-view attention guidance techniques for augmented reality applications and [7] using four different visual effects to guide the user's attention in virtual reality. Both are promising approaches for guiding a user's attention, but not shifting it, to objects 360° around the user.

Audio cues for attention shift

Since one of our two cue designs utilizes both visual and audio representations, we cover a subset of research using these representations. In the work Head-Mounted Sensory Augmentation Device: Comparing Haptic and Audio Modality, the authors Kerdegarhi et al. showed that audio feedback is useful for aiding in indoor navigation [17]. However, the used cues were not multimodal, and therefore lacked the visual part. Löcken et al. showed that participants could react significantly faster to light cues when audio cues were added [19]. In our approach, we chose audio as our second modality because it is already integrated in all head-mounted Virtual and Augmented Reality devices, and therefore requires no additional effort to be implemented. Further, previous work showed it lead to higher reaction times when added to existing visual stimuli.

Cross-modal cueing

From related work, we know about the existence of cross-modal effects [14]. These effects have implications for multi-modal cueing in the endocentric (driver's) reference frame. Many of those effects are related to the combination of visual and auditory stimuli (e.g., the ventriloquism effect [2]). However, many of those effects can be avoided when both cues are presented spatially at the same location and the transported information does not differ between the cues [8].

EYESEE360

In previous work, Gruenefeld et al. developed EyeSee360, a visualization technique for out-of-view objects [10]. The technique is inspired by EdgeRadar [12]. EdgeRadar utilizes the borders of the screen to visualize the positions of off-screen objects. Instead of using screen borders, EyeSee360 uses the periphery of the user. EyeSee360 is shown in Figure 2.

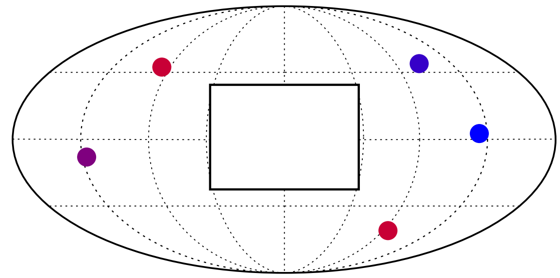


Figure 2: **EyeSee360 technique.** ***Best seen in color.***

The inner square represents the user's FOV. The inner area is not overlapping the user's focus. In this example, EyeSee360 is used in Virtual Reality (VR) glasses. Therefore, the user's FOV is equal to the FOV of the VR screen. When using EyeSee360 in optical see-through Augmented Reality (AR), for example, the inner area will be an ellipse, representing the human field of view. Each dot between the inner area and the outer ellipse represents an out-of-view object, and these dots are called proxies. The color gradient of the proxies from red to blue is encoding how far away the object is from the user (red means close and blue means far away, based on the hot-cold metaphor). The position of the proxy encodes the position of the represented out-of-view object. For example, the blue dot on the right represents an object almost 135 degrees to the right and only a few degrees up. The dotted lines shown in EyeSee360 are helplines: the vertical centered line and the horizontal centered line are considered 0 degree lines. The other dotted lines represent additional 45 degrees relative to the 0 degree lines.

GENERAL APPROACH

To address the problem of shifting the user's attention towards spatially distributed out-of-view objects, we built up on the existing technique EyeSee360 for visualizing these objects. We advance EyeSee360 by exploring different visual cues to shift the user's attention. The attention is shifted towards the representations of out-of-view objects (so called proxies) used by the technique to visualize the positions of these objects. We investigate two different cue designs: (1) non-animated peripheral cues with audio stimuli and (2) animated peripheral cues without audio stimuli. Additionally, we were also concerned with identifying out-of-view objects based on their representations in EyeSee360. To reduce the influence of external factors, we test the different cues in Virtual Reality. Furthermore, current Augmented Reality glasses do not support peripheral vision to the extent required by our implementation.

CUE DESIGN

In this section, we describe our cue design to support shifting the user's attention towards out-of-view objects. We decided to persistently show all relevant out-of-view objects with the visualization technique EyeSee360 to enable users to be aware of surrounding objects at any time. An attention shift is then only required in dangerous encounters. If an out-of-view object becomes dangerous, the proxy is changed into a visual cue towards that object.

Identify out-of-view objects

To evaluate both whether users reacted to the correct cue in our study and whether it would also work to identify out-of-view objects for other cases (e.g., ship docking, where pilots have to be able to distinguish between the different out-of-view objects), we look into how to best identify these out-of-view objects. Identifying specific objects is mostly done by using single or multiple characters, numbers or even combinations of them. This differs from use-case to use-case and can not be described in general. But it is possible to distinguish the lengths of these different encodings used for identifying out-of-view objects. We tried to adapt to these different lengths by using two different kinds of labels (a character and a string). The two different identifications of out-of-view objects are shown as an example in Figure 3. To ensure best readability of these labels we align them towards the center. An example of this alignment can be seen in Figure 6. We chose six characters for the string representation as a typical length of words.

(a) Identification via character.

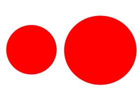
(b) Identification via string.

Figure 3: **Identification of out-of-view objects.**

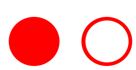
Attention shift with non-animated cue designs

In this paper, we investigate two different cue designs. The first cue design described in this section uses non-animated peripheral cues with audio stimuli. We investigated these cues because Loecken et al. showed that participants could react significantly faster to light cues when audio cues were added in their experiment [19]. However, the presented light cues in their paper were not all presented in the user's periphery. The general idea is that an audio stimulus is given to the user at the same time as the proxy to which the attention should be shifted changes.

A short and smooth change is more likely to stay unrecognized than a fast changing animation [24]. Therefore, we decided to use a short period change for these non-animated cues. In the first step, we identified several possible attributes to be used for the cue design [5]: position, color, size, and shape. Since position and color are already used in EyeSee360, we decided to use size and shape, as seen in Figure 4.



(a) Change of size.

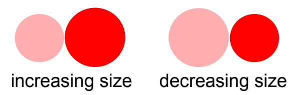


(b) Change of shape.

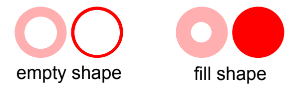
Figure 4: **Non-animated cue designs.**

Attention shift with animated cue designs

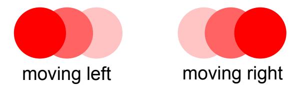
The second cue design described in this section uses animated peripheral cues without audio stimuli. Since motion is well perceivable in the periphery [20], we use an animation inside EyeSee360 as a stimulus to arouse the user's attention. As mentioned before, a longer continuous change is recognized best in the periphery. Therefore, we use kind of a non-stopping pulse animation. We use similar cues as before (but this time with an added pulse animation). Additionally, we use a position animation. The used cues can be seen in Figure 5.



(a) Animation of size.



(b) Animation of shape.



(c) Animation of position.

Figure 5: **Animated cue designs (animation is implemented as non-stopping pulse animation; here from transparent to opaque; then from left to right).**

EXPERIMENT

To evaluate the two different cue designs from the previous section we designed a user study.

Study design

The study was designed as a within-subjects controlled laboratory study. We used Virtual Reality glasses to reduce influencing factors (e.g., lighting conditions). Furthermore, current Augmented Reality glasses do not support peripheral vision to the extent required by our implementation. The study consisted of two counterbalanced parts.

In part A of the study, we investigate if the dependent variable *Reaction time* is influenced by the independent variables *Identification* (Character vs. String) and *Change* (Size vs Shape). This repeated measures within-subjects 2 x 2 factorial design results in four conditions. All conditions were counterbalanced throughout the study.

In part B of the study, we investigate if the dependent variable *Reaction time* is influenced by the independent variables *Identification* (Character vs String) and *Animation* (Size vs. Shape vs. Position). This repeated measures within-subjects 2 x 3 factorial design results in six conditions. All conditions were counterbalanced.

Further, we looked into the error rate for identifying out-of-view objects over both task. Therefore, we compared the two conditions Character and String from the independent variable *Identification*.

We posit the following hypotheses:

H_1 There are fewer errors in identifying out-of-view objects for the character representation compared to the string representation.

H_2 Shifting a user's attention with animated cues is faster than shifting attention with non-animated cues combined with audio stimuli.

H_3 There are significant differences in reaction time between the non-animated cue designs.

H_4 There are significant differences in reaction time between the animated cue designs.

We think that having fewer digits or characters leads to less overlapping, and therefore a lower error rate than a chain of digits or characters would (cp. H_1). Further, we think that an attention shift without audio stimuli and animated cues is faster since it automatically shifts the attention to the correct object, while with audio stimuli the user has to search for the object first (cp. H_2). For shifting the attention with an audio stimuli, a search of the changed proxy is necessary. Therefore, we believe that a larger proxy can be found faster than a proxy that has not changed in size (cp. H_3). For shifting the attention with animated cues, we think there is no difference in reaction time because we think the kind of animation is not influencing users' reaction times (cp. H_4).

Implementation

The source code of EyeSee360 has been made available by the authors under MIT-License on Github[1]. EyeSee360 is implemented using Unity and therefore, supports various platforms (e.g., Hololens, Oculus, etc.). We advanced this code to support an identification based on characters and strings. Further, we added support for the two different kinds of visual cues. An example of this implementation can be seen in Figure 6.

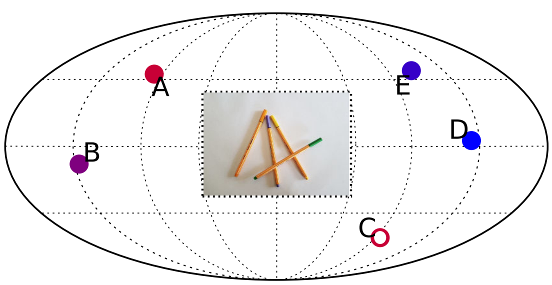


Figure 6: **Implementation of non-animated cues with Character and Shape as cue design. *Best seen in color.***

Apparatus

During the study, participants were equipped with an Oculus Rift, a Virtual Reality headset with earphones included. Participants were seated on a chair in front of a table and they were given a keypad as an input device to measure their reaction times. Figure 6 shows the VR view. The proxies are in different colors because they represent the distances towards the out-of-view objects (from red/close to blue/far away).

Procedure

In the beginning participants were introduced to EyeSee360. An example scene with several out-of-view objects was shown to the participant. Afterwards, participants started with part A or B, depending on the counterbalanced design. Both parts had the same primary task. The primary task consisted of a slide show of images (see Figure 7 for examples). Each image contained a random number of random objects (between three to eight). The participant had to count these objects and state the number of objects seen. The secondary task then was dependent on which part the participant was currently working. Each part was introduced by a demo showing each cue once. Each cue was tested six times[2].

(a) Four pens. (b) Five arrows. (c) Three bonbons.

Figure 7: **Examples for primary task.**

Part A: Non-animated cues Each non-animated cue together with an *Identification* was tested in one block. While participants were focusing on the primary task, at a random point in time (between 1 to 10 seconds) one of the proxies was changed using the cue of the current block, and at the same time a sound stimuli was presented to the participant. Participants were asked to react to the sound stimuli by searching for the changed proxy. After finding it, they had to press the key on the keypad. Afterwards, they had to identify the object by saying out loud the character or string. They could proceed with the next trial by pressing the key on the keypad.

Part B: Animated cues In this part, each animated cue was tested in one block together with each *Identification*. While participants were focusing on the primary task, at a random point in time (between 1 to 10 seconds) one of the proxies was animated by the cue of the current block. Participants were asked to react to these changes in their peripheries by locating the animated proxies and pressing a key. After they pressed the key, they had to identify the object by saying out loud the character or string. The next trial started by pressing any key.

After each part, participants filled out a questionnaire. They had to rate each cue design on a 6-point Likert item scale. Additionally, they were asked for the cues they found best and worse. At the end, they could give further comments.

Participants

We recruited 18 participants (6 female), aged between 22 and 49 years (M=27.72, SD=7.98). None suffered from color vision impairments, 13 had normal vision and 5 had corrected to normal vision. Five participants had no experience with HMDs, 4 were somewhat familiar with such devices, 2 had some experience, 6 were experienced, and 1 participant was very experienced (Md=2.50, IQR=1.25-4).

Results

Identification error rate For trials in which String was used as a label, participants had problems identifing the out-of-view objects in 37 of 540 cases (6.85%). For trials in which Character was used as a label, participants had problems identifing the out-of-view objects in 19 of 540 cases (3.52%). The values can be seen in Table 1. In summary, String has a higher error rate (6.85%) than Character (3.52%). We compared both values using a chi-squared test as suggested by Campbell et al.

[1] https://github.com/UweGruenefeld/EyeSee, last retrieved July 21, 2018

[2] Derived as an optimal number of iterations from pretesting.

[16] ($\chi^2(1)$=6.08, p=0.014, N=18). Here, we got a significant result and thus we can accept our hypothesis H_1.

Condition	total	failed	hesitated	combined
String	540	13 (2.41%)	24 (4.44%)	37 (6.85%)
Character	540	5 (0.93%)	14 (2.59%)	19 (3.52%)

Table 1: **Comparison of identification error rates for both conditions (failed means that participants were not able to identify the out-of-view objects and hesitated means that participants were hesitating and not directly answering).**

Animated vs. non-animated cues We compare the reaction times of both parts (mean: non-animated=1.10s and animated=0.86s, median: non-animated=0.74s and animated=0.66s) to test our second hypothesis H_2. To compare the reaction times we had to exclude the animation of position from the animated cues to have two equal sized groups. Here, we compared change of size and change of shape vs. animation of size and animation of shape. A Shapiro-Wilk-Test showed that our data is not normally distributed (p<0.001). As we compare two matched groups within subjects, we directly performed a Wilcoxon Signed-rank test. Here we found a significant effect of task (animated vs. non-animated) on reaction time (W=66000, Z=7.41, p<0.001, ϕ=0.25). Therefore, we can accept H_2.

Reaction time of non-animated cues Here, we consider the effects of the Combination (*Identification* and *Change*) on mean reaction time (in seconds). The mean reaction times are: Size+String=1.24s, Size+Character=1.04s, Shape+String=0.97s and Shape+Character=1.13s. The reaction times are compared in Figure 8. A Shapiro-Wilk-Test showed that our data is not normally distributed (p<0.001), and thereafter we ran a Friedman test that revealed no significant effect of *Identification* and *Change* on reaction times ($\chi^2(3)$=4.77, p=0.19, N=18). Therefore, we can not accept our third hypothesis H_3.

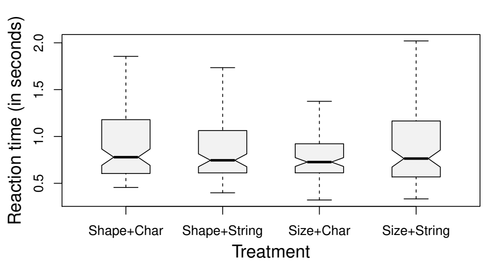


Figure 8: **Median time for non-animated cues.**

Subjective measurement of non-animated cues To subjectively evaluate the combinations of *Identification* and *Change*, we asked participants to rate them on a 6-point Likert item questionnaire (where 6 is very good and 1 is very bad). The results are Size+String (Md=5, IQR=4-6), Size+Character (Md=4, IQR=4-5), Shape+String (Md=5, IQR=5-6) and Shape+Character (Md=5, IQR=4-5). We ran a Friedman test that revealed no significant differences between the tested

Comparision	p-value	ϕ-value
Size+String vs. Size+Character	< 0.01	0.46
Size+String vs. Shape+Character	< 0.01	0.50
Shape+Character vs. Position+String	< 0.01	0.54
Shape+Character vs. Position+Character	< 0.01	0.51

Table 2: **Comparison for subjective measurement of animated cues (contains only significant results).**

combination ($\chi^2(3)$=6.72, p=0.08, N=18).

Reaction time of animated cues We consider the effects of two factors (*Identification* and *Animation*) on mean reaction time (in seconds). The mean reaction times are: Size+String=1.24s, Size+Character=1.04s, Shape+String=0.97s, Shape+Character=1.13s, Position+String=1.02s and Position+Char=0.94s. The reaction times are compared in Figure 9. A Shapiro-Wilk-Test showed

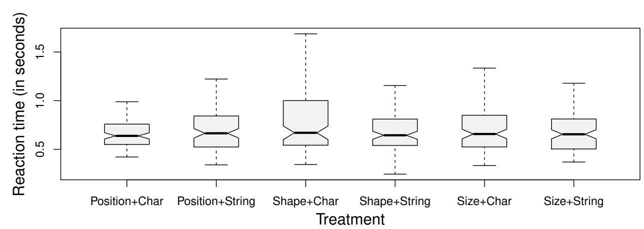


Figure 9: **Median time for animated cues.**

that our data is not normally distributed ($p < 0.001$), and thereafter we ran a Friedman test that revealed no significant effect of Identification and Animation on reaction times ($\chi^2(5)$=2.17, p=0.8245, N=18). Therefore, we can not accept our hypothesis H_4.

Subjective measurement of animated cues To subjectively evaluate the combinations of *Identification* and *Animation* we asked participants to rate them on a 6-point Likert item questionnaire (where 6 is very good and 1 is very bad). The results are Size+String (Md=5, IQR=4.25-5.75), Size+Character (Md=4, IQR=3.25-5), Shape+String (Md=5, IQR=4-5), Shape+Character (Md=4, IQR=3-4), Position+String (Md=6, IQR=5-6) and Position+Character (Md=5, IQR=3.25-5.75). Therefore, Position+String is subjectively perceived best. We ran a Friedman test that revealed significant differences between the tested combinations ($\chi^2(5)$=27.49, $p < 0.001$, N=18). A post-hoc test using Wilcoxon Signed-rank with Bonferroni correction showed significant differences of some conditions (see Table 2).

Discussion

Reaction times There were no significant differences in reaction time between the treatments for non-animated cues as well as for animated cues. Therefore, we can not accept our hypotheses H_3. We argue that if attention is triggered visually at the same position in the periphery, it does not matter how. Rather than investigating more cue designs, further studies should look into the effects of different cue positions in the periphery. It is possible that cues presented more distantly in the periphery are perceived with different reaction times than cues presented in the near periphery [20].

Subjective performance Even if there were no significant differences in reaction times between the treatments, participants subjectively preferred specific cues over others. For non-animated cues, Shape+String was perceived subjectively best, while Position+String was subjectively perceived best for animated cues. However, most important is the fact that participants preferred String over Character.

Error rate As stated in hypothesis H_1, the error rate was lower with character compared to string. This is supported by the statements from the participants. However, participants preferred String over Character. The higher error rate was due to more frequent overlapping for String.

Ecological validity In this study we tested a Virtual Reality setup. We did this to exclude external parameters and focus on gaining a better understanding of shifting user's attention with peripheral cues. Furthermore, optical see-through Augmented Reality is influenced by light and lacks technical solutions that fully support peripheral visualizations required by our approach. However, we believe that our lab study is promising and extendable in future work with head-mounted Augmented Reality, in a driving simulator for example. Further, it might also be interesting for other use-cases, such as being aware of one's surroundings during highly automated driving [3]. The primary task in this study was artificial and had limited connection to the more complex tracking and object-and-event activities associated with driving. Future research should look at more realistic primary tasks (e.g., driving and overtaking on a highway).

Study limitations Shifting the attention onto the proxies in EyeSee360 is not enough to ensure that users are aware of objects such as approaching cars. However, in previous studies, Gruenefeld et al. investigated search time performance and out-of-view objects awareness in EyeSee360 [10]. We argue that our results are connectable to previous research because when users are able to locate out-of-view objects by looking at the proxies, they should also be able to do so in our scenario.

CONCLUSION AND FUTURE WORK

In this work, we investigated ways to shift a user's attention to spatially distributed out-of-view objects. We identified EyeSee360 as a suitable visualization technique and explored different peripheral cues in a controlled lab study. Our results showed that animated peripheral cues can shift attention in under one second on average. For shifting the attention with non-animated cues, it takes an average of 1.1 seconds. Further, we investigated different approaches to identify out-of-view objects. We showed that users prefer identification via strings. However, this leads to more clutter and a higher error rate. Our work may be a useful precursor to further studies of cue designs and positions in an automotive simulator with Virtual Reality. Furthermore, future work should investigate our approach with head-mounted Augmented Reality in a driving simulator.

REFERENCES

1. Patrick Baudisch and Ruth Rosenholtz. 2003. Halo: A Technique for Visualizing Off-screen Objects. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03)*. ACM, New York, NY, USA, 481–488. DOI: http://dx.doi.org/10.1145/642611.642695

2. Paul Bertelson. 1999. Chapter 14 Ventriloquism: A case of crossmodal perceptual grouping. In *Cognitive Contributions to the Perception of Spatial and Temporal Events*, Gisa Aschersleben, Talis Bachmann, and Jochen Müsseler (Eds.). Advances in Psychology, Vol. 129. North-Holland, 347 – 362. DOI: http://dx.doi.org/10.1016/S0166-4115(99)80034-X

3. Shadan Sadeghian Borojeni, Lewis Chuang, Wilko Heuten, and Susanne Boll. 2016. Assisting Drivers with Ambient Take-Over Requests in Highly Automated Driving. In *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (Automotive'UI 16)*. ACM, New York, NY, USA, 237–244. DOI: http://dx.doi.org/10.1145/3003715.3005409

4. Stefano Burigat, Luca Chittaro, and Silvia Gabrielli. 2006. Visualizing Locations of Off-screen Objects on Mobile Devices: A Comparative Evaluation of Three Approaches. In *Proceedings of the 8th Conference on Human-computer Interaction with Mobile Devices and Services (MobileHCI '06)*. ACM, New York, NY, USA, 239–246. DOI: http://dx.doi.org/10.1145/1152215.1152266

5. William S. Cleveland and Robert McGill. 1984. Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *J. Amer. Statist. Assoc.* 79, 387 (1984), pp. 531–554.

6. Andy Cockburn, Amy Karlson, and Benjamin B. Bederson. 2009. A Review of Overview+Detail, Zooming, and Focus+Context Interfaces. *ACM Comput. Surv.* 41, 1, Article 2 (Jan. 2009), 31 pages. DOI: http://dx.doi.org/10.1145/1456650.1456652

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

8. Jon Driver and Charles Spence. 2000. Multisensory perception: Beyond modularity and convergence. *Current Biology* 10, 20 (2000), R731 – R735. DOI: http://dx.doi.org/10.1016/S0960-9822(00)00740-5

9. Uwe Gruenefeld, Abdallah El Ali, Wilko Heuten, and Susanne Boll. 2017a. Visualizing Out-of-view Objects in Head-mounted Augmented Reality. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '17)*. ACM, New York, NY, USA, Article 81, 7 pages. DOI: http://dx.doi.org/10.1145/3098279.3122124

10. Uwe Gruenefeld, Dag Ennenga, Abdallah El Ali, Wilko Heuten, and Susanne Boll. 2017b. EyeSee360: Designing a Visualization Technique for Out-of-view Objects in Head-mounted Augmented Reality. In *Proceedings of the 5th Symposium on Spatial User Interaction (SUI '17)*. ACM, New York, NY, USA, 109–118. DOI: http://dx.doi.org/10.1145/3131277.3132175

11. Sean Gustafson, Patrick Baudisch, Carl Gutwin, and Pourang Irani. 2008. Wedge: Clutter-free Visualization of Off-screen Locations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*. ACM, New York, NY, USA, 787–796. DOI: http://dx.doi.org/10.1145/1357054.1357179

12. Sean G. Gustafson and Pourang P. Irani. 2007. Comparing Visualizations for Tracking Off-screen Moving Targets. In *CHI '07 Extended Abstracts on Human Factors in Computing Systems (CHI EA '07)*. ACM, New York, NY, USA, 2399–2404. DOI: http://dx.doi.org/10.1145/1240866.1241014

13. Niels Henze, Benjamin Poppinga, and Susanne Boll. 2010. Experiments in the Wild: Public Evaluation of Off-screen Visualizations in the Android Market. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries (NordiCHI '10)*. ACM, New York, NY, USA, 675–678. DOI:http://dx.doi.org/10.1145/1868914.1869002

14. Cristy Ho and Charles Spence. 2017. *The multisensory driver: Implications for ergonomic car interface design*. CRC Press.

15. Liqiang Huang, Anne Treisman, and Harold Pashler. 2007. Characterizing the limits of human visual awareness. *Science* 317, 5839 (2007), 823–825.

16. Campbell Ian. 2007. Chi-squared and Fisher-Irwin tests of two-by-two tables with small sample recommendations. *Statistics in Medicine* 26, 19 (2007), 3661–3675. DOI:http://dx.doi.org/10.1002/sim.2832

17. Hamideh Kerdegari, Yeongmi Kim, and Tony J. Prescott. 2016. Head-Mounted Sensory Augmentation Device: Comparing Haptic and Audio Modality. In *Biomimetic and Biohybrid Systems*, Nathan F. Lepora, Anna Mura, Michael Mangan, Paul F.M.J. Verschure, Marc Desmulliez, and Tony J. Prescott (Eds.). Springer International Publishing, Cham, 107–118.

18. Yen-Chen Lin, Yung-Ju Chang, Hou-Ning Hu, Hsien-Tzu Cheng, Chi-Wen Huang, and Min Sun. 2017. Tell Me Where to Look: Investigating Ways for Assisting Focus in 360 Video. In *Proc. CHI '17*. ACM, New York, NY, USA, 2535–2545. DOI: http://dx.doi.org/10.1145/3025453.3025757

19. Andreas Löcken, Sarah Blum, Tim Claudius Stratmann, Uwe Gruenefeld, Wilko Heuten, Susanne Boll, and Steven van de Par. 2017. Effects of Location and Fade-in Time of (Audio-)Visual Cues on Response Times and Success-rates in a Dual-task Experiment. In *Proceedings of the ACM Symposium on Applied Perception (SAP '17)*. ACM, New York, NY, USA, Article 12, 4 pages. DOI: http://dx.doi.org/10.1145/3119881.3119894

20. Suzanne P. Mckee and Ken Nakayama. 1984. The detection of motion in the peripheral visual field. *Vision Research* 24, 1 (1984), 25 – 32. DOI: http://dx.doi.org/10.1016/0042-6989(84)90140-8

21. Marie-Christin Ostendorp, Jan Charles Lenk, and Andreas Lüdtke. 2015. Smart Glasses to Support Maritime Pilots in Harbor Maneuvers. *Procedia Manufacturing* 3 (2015), 2840 – 2847. DOI: http://dx.doi.org/10.1016/j.promfg.2015.07.775

22. Michael I. Posner and Steven E. Petersen. 1990. The Attention System of the Human Brain. *Annual Review of Neuroscience* 13, 1 (1990), 25–42. DOI: http://dx.doi.org/10.1146/annurev.ne.13.030190.000325

23. P. Renner and T. Pfeiffer. 2017. Attention guiding techniques using peripheral vision and eye tracking for feedback in augmented-reality-based assistance systems. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*. 186–194. DOI: http://dx.doi.org/10.1109/3DUI.2017.7893338

24. Daniel J. Simons and Daniel T. Levin. 1997. Change blindness. *Trends in Cognitive Sciences* 1, 7 (1997), 261 – 267. DOI: http://dx.doi.org/10.1016/S1364-6613(97)01080-2

25. Marcus Tönnis, Christian Sandor, Christian Lange, and Heiner Bubb. 2005. Experimental Evaluation of an Augmented Reality Visualization for Directing a Car Driver's Attention. In *Proceedings of the 4th IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR '05)*. IEEE Computer Society, Washington, DC, USA, 56–59. DOI: http://dx.doi.org/10.1109/ISMAR.2005.31

26. Steven Yantis. 2000. Goal-Directed and Stimulus-Driven Determinants of Attentional Control. (2000).

27. Polle T. Zellweger, Jock D. Mackinlay, Lance Good, Mark Stefik, and Patrick Baudisch. 2003. City Lights: Contextual Views in Minimal Space. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems (CHI EA '03)*. ACM, New York, NY, USA, 838–839. DOI:http://dx.doi.org/10.1145/765891.766022