Towards capturing focal/ambient attention during dynamic wayfinding

JAKUB KRUKAR, Institute of Geoinformatics, University of Muenster, Germany
PANAGIOTIS MAVROS, ETH Zürich, Future Cities Laboratory, Singapore ETH Centre, Singapore
CHRISTOPH HÖLSCHER, Chair of Cognitive Science, Department of Social Sciences, ETH Zürich, Switzerland

This work-in-progress paper reports on an ongoing experiment in which mobile eye-tracking is used to evaluate different wayfinding support systems. Specifically, it tackles the problem of detecting and isolating attentional demands of building layouts and signage systems in wayfinding tasks. The coefficient K has been previously established as a measure of focal/ambient attention for eye-tracking data. Here, we propose a novel method to compute coefficient K using eye-tracking from virtual reality experiments. We detail challenges associated with transforming a two-dimensional coefficient K concept to three-dimensional data, and the debatable theoretical equivalence of the concept after such a transformation. We present a preliminary implementation to experimental data and explore the possibilities of the method for novel insight in architectural analyses.

Additional Key Words and Phrases: virtual reality; wayfinding; eye-tracking; usability; attention

ACM Reference Format:
Jakub Krukar, Panagiotis Mavros, and Christoph Hölscher. 2020. Towards capturing focal/ambient attention during dynamic wayfinding. In *Symposium on Eye Tracking Research and Applications (ETRA '20 Adjunct), June 2–5, 2020, Stuttgart, Germany*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3379157.3391417

1 INTRODUCTION

Architects are interested in understanding variations in the visual attention of building users caused by the building's layout [Kaicker et al. 2019; Zook and Bafna 2016]. To date, modelling the influence of building layouts on visual attention has been achieved primarily by using isovist-based visibility models (i.e., calculations of visibility relations between possible locations of building users and different parts of the building) [Benedikt 1979; Derix et al. 2008; Turner et al. 2001], and validating the cognitive adequacy of these models in separate laboratory-based experiments [Krukar and Conroy Dalton 2013; Lu and Ye 2019].

The combination of mobile eye-tracking and virtual reality offers a new possibility for studying the effect of building layout variations on attention: by recording visual attention during realistic tasks in virtual models of yet-unbuilt layouts [Kuliga et al. 2015].

However, traditional eye-tracking metrics are not well-suited to assist in typical architectural design problems because they focus on the target of the gaze: gaze is used to find out what objects were looked at, and resulting metrics describe some property of the gaze target. Architectural analyses, however, typically focus on the origin of the gaze (i.e., locations of building users) and are specifically interested in modelling properties of building locations from which gaze originated. These are different issues because the former approach might aggregate information per gaze target, regardless of where gaze originated from; while the latter approach might aggregate information per building location, regardless of what was the target of the gaze. For example, attention maps [Schrom-Feiertag et al. 2017] can identify signage that is seen or ignored by navigators but does not provide insights into building areas *at which* users were confused. Landmark studies [Ohm et al. 2017, 2014; Wenczel et al. 2017] also analyse which objects are (or are not) fixated, but do not formally analyse the gaze origin.

Previous research demonstrated that this is a significant limitation. For example, parts of the viewing field that users fixate before making navigational decisions differ, depending on spatial properties of the location from which the decision is being made [Emo 2018; Wiener et al. 2012]. The angle of approach towards architecturally important objects and signs also makes a difference to navigational behaviour [Müller-Feldmeth et al. 2014]. Neither of these issues can be captured by metrics that do not take into consideration the location from which gaze originated.

We present a work-in-progress implementation of the Coefficient K [Krejtz et al. 2016]–an eye-tracking measure of focal-vs-ambient attention mode [Velichkovsky et al. 2005]–to a dynamic navigational scenario in a virtual reality building simulation.

Coefficient K is a measure of visual behaviour fluctuating between focal and ambient mode of viewing. Focal mode is operationalised by the presence of relative long fixations followed by relatively small saccades. Ambient mode is operationalised by the presence of relatively short fixations followed by relatively large saccades. Focal attention has been associated with deeper cognitive processing of stimuli while ambient attention has been associated with scanning or exploring the stimuli. It has been evaluated in viewing artworks [Krejtz et al. 2016] and maps [Krejtz et al. 2017]. K is derived by calculating the difference between fixation duration and its subsequent saccade expressed in standardised (z-score) values. Negative K indicates the ambient mode of viewing, positive K indicates the focal mode of viewing; and K equal or close to 0 indicates that a fixation of an average duration has been followed by a saccade of an average amplitude.

Coefficient K can be a useful measurement unit for architectural design evaluation because:

(1) Unlike traditional eye-tracking measures (e.g., fixations durations), focal/ambient attention mode is directly interpretable as a property of the viewer (gaze origin), and not the property of the gaze target.
(2) It indexes the *mode*, and not the *quantity* of attention. This might indicate a changing architectural experience even when

ETRA '20 Adjunct, June 2–5, 2020, Stuttgart, Germany

ACM ISBN 978-1-4503-7135-3/20/06.
https://doi.org/10.1145/3379157.3391417

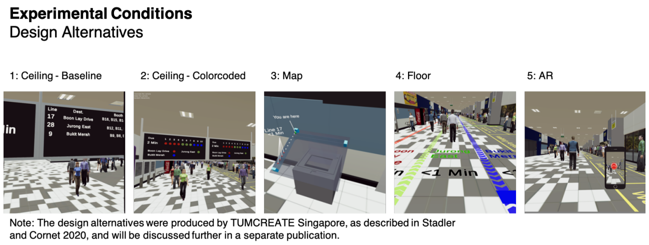

Fig. 1. Screenshots of the five (5) experimental conditions

the cumulative *quantity* of attention dedicated to the surrounding objects remains unchanged [Krukar and Dalton 2020].

(3) It can be indexed over time and can therefore indicate dynamic fluctuations in attention modes as navigators travel through a building.

The paper reports on preliminary results of a virtual reality experiment showing the possible use of coefficient K in a pre-occupancy evaluation scenario. Specifically, this paper identifies challenges arising from applying a two-dimensional concept of coefficient K to three-dimensional data, points at possible extensions of coefficient K that may solve the newly revealed issues, and provides ideas for future work.

2 EXPERIMENT

The experiment design was motivated by the challenge of evaluating different proposals for wayfinding signage in the context of a large public transport hub.

2.1 Participants

Participants were recruited from the student pool of local universities in Singapore, and were compensated for their time. The study was approved by the ETH Zürich Ethics Committee (B_EK_2019-N111). We report analyses from an ongoing data collection for a subset of 11 participants.

2.2 Experimental design

Each participants completed three (3) wayfinding tasks in each of five (5) conditions (Fig. 1) corresponding to different signage design alternatives. Therefore, in total, each participant completed fifteen (15) tasks (within-subjects design).

2.3 Procedure

Participants were received by the experimenter and informed consent was obtained. The primary task was explained to them and they were given a demonstration of how to use the interface to navigate within the virtual environment. Navigation was performed by using the arm-swing method. Users hold an HTC VIVE controller with each hand, and move their hands imitating arm movement during walking. Then, the head-mounted display with a built-in eye-tracker was placed and adjusted on their head, and they completed the eye-tracking calibration. Subsequently, participants completed a total

Fig. 2. Left: A photograph of the Bus Interchange that served as the basis for this study. Right: Screenshot from the virtual environment, showing the gaze direction (note that the yellow line was not visible to participants).

of fifteen (15) wayfinding tasks: three (3) in each of five (5) experimental conditions. The tasks emulated the experience of boarding a bus at a major transport interchange. Participants either started from one of the three entrances of the Bus Interchange, or arrived inside a bus (to emulate to the experience of transit from one bus to another). At the beginning of each task, participants were given a naturalistic destination (i.e. a place in Singapore, such as Jurong East) and their goal was to find the berth (gate) and board in the next bus that goes in that destination. After completion of all wayfinding tasks, participants completed a questionnaire about their experience. Each task took M = 101 seconds (SD = 54s) on average.

2.4 Materials

The experiment was conducted in a virtual replica of an existing Bus Interchange in Singapore. The Bus Interchange consists of two major hallways, measuring 150 by 12 meters and 160 by 12 meters respectively, with a total surface of 1800 sq.m. The hallways are arranged perpendicular to each other, forming a L-shape. A series of berths of boarding buses (gates) are positioned along three of the long sides of the hallways. For the virtual experiments, only one group of berths was used.

Five different wayfinding aids were developed by [Stadler and Cornet 2020] to guide passengers to a specific berth. These included 2 types of ceiling-mounted displays, 1 type of a handheld augmented reality device, 1 floor-based directional system, and 1 map (Figure 1).

2.5 Apparatus

The virtual scene was rendered in real-time using the Unity 3D game engine (version 2019.1.3f1), on a desktop computer (equipped with Intel Core i-7 8700K @3.7GHz, 16GB RAM, and NVIDIA GeForce GTX 1080 graphics card).

For the presentation of materials and the recording of eye-tracking, an HTC Vive Pro Eye head-mounted display was used. The device is equipped with dual AMOLED screens providing a resolution of 1440 x 1600 pixels per eye with a field of view of 110 degrees and a refresh rate of 90Hz. Eye-tracking data were obtained by a Tobii Eye-tracking module which is embedded in the HTC VIVE Pro EYE HMD. It operates at a manufacturer-reported sampling frequency of 120Hz with an accuracy of 0.5 - 1.1 degrees. The eye-tracking data are registered using the HTC SRanipal SDK, which provides the timestamp, the location of gaze origin, and the location of gaze target.

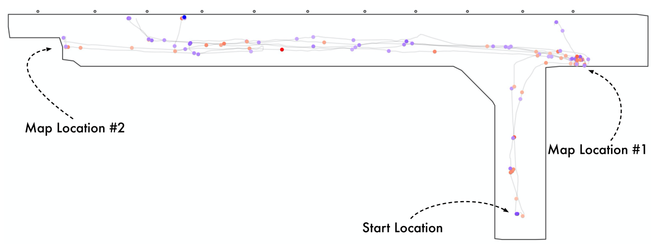

Fig. 3. Example of a wayfinding trajectory and gaze behaviour of one participant in condition 3: Map. Medium values (i.e. K in the range -1 to 1) are not displayed for the clarity of the visualisation. Red indicates focal attention mode, and blue indicates ambient attention mode.

2.6 Extracting coefficient K for three-dimensional data

We applied a dispersion-based algorithm to detect fixations with a threshold of 4 degrees. This was calculated based on the angle between two vectors in the three-dimensional space: one vector defined by the gaze origin (i.e., the position of the camera in the virtual model) and the target of the current gaze; the second vector defined by the gaze origin and the target of the subsequent gaze. We then removed fixations shorter than $50ms$, longer than $3000ms$, and we treated all periods longer than $500ms$ between two subsequent fixations as a period of missing data (as opposed to treating it as a saccade).

We calculated K for each fixation based on the approach of [Krejtz et al. 2016]. Because the output provided by the SRanipal SDK is pre-filtered, we only extracted K for those data, where the location and duration of two subsequent fixations could be reliably estimated. Consequently, the analyses here reported involve missing data, but the presented analytical approach would remain unchanged if a more reliable dataset was available.

3 PRELIMINARY RESULTS AND DISCUSSION

We detected, on average, 119 fixations per participant, per task, with an average fixation length of $M = 430ms$ ($SD = 102ms$). Average angular distance of saccades (only those that occurred between two reliably detected fixations) was $M = 14.4°$($SD = 5.8°$).

We propose two approaches to considering coefficient K for the architectural analysis: a single-participant analysis (Fig. 3) and an aggregated between-participants analysis (Fig. 4). The single-participant analysis (Fig. 3) can be used to identify episodes of focal attention near the start and along the participant's trajectory. Episodes of ambient attention can be observed near the map locations, where the participant retrieved information, as well as during linear trajectories where the participant walked in a single direction. We hypothesise that the combination of ambient and focal attention near map locations might result from distinct processes of analysing the map, and comparing it to the surrounding environment.

In the between-participants analysis (Fig. 4), areas of blue/red might be indicative of a common visual experience across participants in these building locations. Note that the values correspond to a property of the gaze origin (attentional mode of participants whilst traversing this location) and not the property of the gaze target.

4 DISCUSSION AND OPEN CHALLENGES

The presented results are inconclusive. This could potentially be the result of an interaction between the building typology and the nature of the task. The building provides a large open space (12 meters width) with long lines of sight in all directions and there are few spatial decision points. The task requires people to quickly scan the environment to identify the correct berth (gate number) once, with lower attentional demands for the remainder of the trial. A more complex building with many corridors and intersection would be more likely to evoke episodes of ambient and focal attention that are spatially determined. We are collecting more data to determine that.

So far, in our application, the coefficent K seems to be inconsistent with its theoretical explanations. In the single-participant analysis, it is surprising that episodes of ambient attention are mixed with episodes of focal attention along the entire trajectory. Based on previous research with stationary eye-tracking [Krejtz et al. 2017] it could be rather expected that the attentional mode changes gradually over time, and in addition that it changes swiftly with the task at hand (e.g., reading a map vs. walking along a corridor). In the between-participants analysis, clusters of focal attention could be expected to be larger than single 2x2 meter grid cells. The location and amount of these clusters is also surprisingly similar across different conditions that should rather have a significant effect on participants' attention.

In this section we review four open challenges that might underlie the inconsistency between our results, and the theoretical explanation of the coefficient K. Data gathered in our ongoing experiment is a comprehensive benchmark for the application of coefficient K to dynamic wayfinding experiments, because it includes a large number of conditions that affect attention, a combination of multiple cognitive tasks within each single condition (e.g., identifying a map, reading a map, and navigating through architectural space), as well as multiple data collected from single participants (that allows for a better control of individual differences). The present challenges are inspired by the application of coefficient K to such a diverse dataset and therefore focus on issues that are likely to distinguish a stationary application of coefficient K, from its application in dynamic wayfinding experiments.

4.1 Challenge 1: Theoretical equivalence of the two-dimensional and three-dimensional coefficient K

It is possible to study a map located in the environment with either focal or ambient mode of attention [Krejtz et al. 2017], but is it possible (within the theoretical definition of the concept) to look at large architectural elements with either focal or ambient mode of attention? With the current implementation, the detection of ambient/focal attention modes might be inconsistent across these two situations.

Originally, focal/ambient attention modes were defined for the task of engaging a two-dimensional stimuli represented in front of a participant on a computer screen. This arrangement restricts possible saccade amplitudes, since the entire stimuli is present within a relatively narrow sub-part of the visual field. In a dynamic navigational scenario, with eye movement recorded by a head-mounted

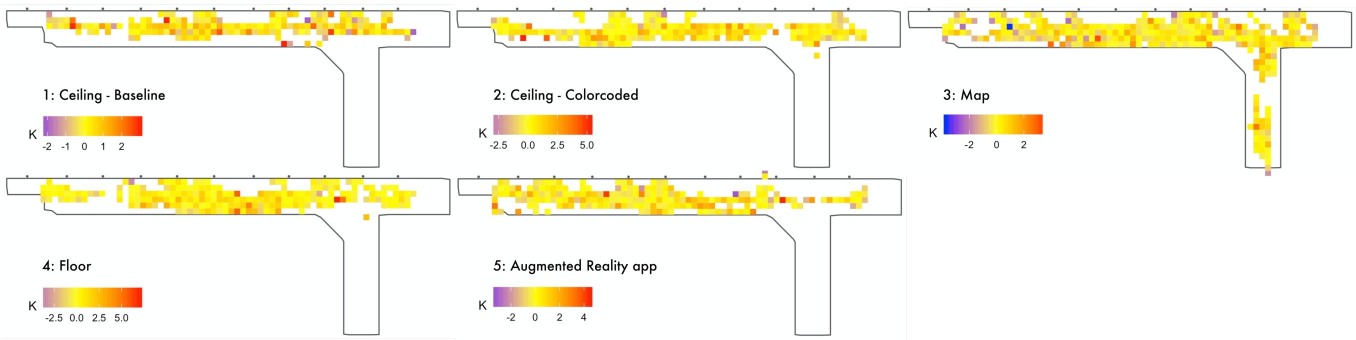

Fig. 4. Heatmaps showing the average (mean) coefficient K of multiple participants in a 2 X 2 meter raster grid across five experimental conditions.

mobile eye-tracker, the possible saccade amplitude is much larger. For example, a person fixating a sign in front of them, with rapid head rotation might subsequently fixate a sign at a 90-degree angle from it.

The presence of so large saccades makes all saccades occurring within single, smaller objects (e.g., signs) relatively small (in terms of their z-scores). Thus, the current approach might be biased, by missing the ambient mode of attention occurring within smaller objects, or across objects located close to each other.

This poses the question whether all saccades in dynamic navigational scenarios should be considered as the same category of events. One solution to the described problem could be to calculate focal/ambient attention mode separately within and across objects of interest, or to weight saccade amplitudes before they are submitted to the coefficient K formula.

4.2 Challenge 2: The choice of filtering out visual engagement with wayfinding aids

Dynamic navigational experiments typically include external wayfinding aids that need to be consulted by the participant in order to inform their navigational behaviour. This might introduce a bias to the eye movement data used for the coefficient K calculation, as there is an unusually high number of fixations on a single object, that are likely to follow a homogenous pattern (e.g., a handheld device is always located at a similar distance, is likely to be engaged with similar fixation durations, and the saccades following this engagement are likely to be falling onto the environment directly in front of the walking participant). This poses the question whether some interactions should be explicitly filtered out from the calculations, as they introduce a large amount of homogenous data points that affect the normalization of saccade lengths and fixation durations in the coefficient K formula. Simultaneously, these interactions are not informative about the engagement with the environment.

4.3 Challenge 3: Accounting for different types of attention objects

Initially, coefficient K was developed to investigate visual interaction with stimuli displayed on a 2-dimensional plane such as a screen. In such scenarios it can be assumed that all stimuli are task-related and thus the gaze behaviour can reveal different attentional processes.

On the contrary, in real buildings, as well as in populated virtual environments (as in the present study), people routinely distribute their attention between the physical environment (floors, walls, doors, signage, etc) and the social environment (i.e. other people). In this context, the saccadic behaviour underlying coefficient K is the result of distinct motivations and cognitive processes. This is especially the case for high values of K (ambient attention), the product of large saccades, which may be due to seeking information (e.g. looking for a sign), looking at other pedestrians out of interest, guiding locomotion-related decision making such as collision avoidance, or due to distraction.

4.4 Challenge 4: Normalising coefficient K across experimental trials

In its original formalisation, coefficient K uses fixation durations and saccade amplitudes normalised across all experimental tasks and conditions, in order to reveal true between-condition differences. However, differences across subsequent stimuli presented on a computer screen are likely to be smaller compared to differences in eye movement imposed by different building layouts, or navigational tasks. For example, asking participants to perform a back-tracking navigational tasks might expand the range of saccade amplitudes, since participants will start to look back as they walk through the environment. This expanded range of saccade amplitudes will affect the z-scores of saccades in non-backtracking tasks (where participants restrict themselves to looking ahead).

5 CONCLUSION

Coefficient K is a promising high-level metric of visual behaviour that has the potential for improving architectural analysis and introducing a new formalised method for studying the impact of building layouts on users' visual experience and behaviour. However, our implementation revealed that a direct application of coefficient K to a dataset from a dynamic wayfinding experiment introduces novel challenges. We identified four open challenges that will form the basis for our future work in this area.

ACKNOWLEDGMENTS

The authors wish to thank Dr Henriette Cornet and Sebastian Stadler for the VR materials, Zaol-Kefli Muhammad Salihin for implementing the experiment in Virtual Reality, and Angel Yuen for conducting the data collection.

Funding. JK acknowledges the support of the European Research Council, Grant Agreement No 637645 and Future Cities Laboratory for a visiting fellowship grant. PM and CH acknowledge funding support for the project "From Virtual Reality to Simulation: User-Centred Design of Dynamic Guidance Systems for transit hubs" from the National Research Foundation, Prime Minister's Office, Singapore, under its Campus for Research Excellence and Technological Enterprise (CREATE) programme, Grant No NRF2018-ITS003-015. CH and PM also acknowledge research conducted at the Future Cities Laboratory at the Singapore-ETH Centre, which was established collaboratively between ETH Zurich and Singapore's National Research Foundation (FI 370074016) under its Campus for Research Excellence and Technological Enterprise programme.

REFERENCES

M.L. Benedikt. 1979. To take hold of space: isovists and isovist fields. *Environment and Planning B: Planning and Design* 6 (1979), 47–65.

Christian Derix, Åsmund Gamlesæter, and Pablo Miranda Carranza. 2008. 3d Isovists and Spatial Sensations: Two methods and a case study. In *Movement and Orientation in Built Environments: Evaluating Design Rationale and User Cognition*, Saif Haq, Christoph Hölscher, and Sue Torgrude (Eds.). Universitat Bremen, Bremen, 67–72.

Beatrix Emo. 2018. Choice zones: architecturally relevant areas of interest. *Spatial Cognition & Computation* 18, 3 (jul 2018), 173–193. https://doi.org/10.1080/13875868.2017.1412443

Arjun Kaicker, Uli Blum, Philipp Siedler, and Lorena Espaillat. 2019. Enhancing Workplace Design through Advanced Floor Plate Analytics. *Technology Architecture and Design* 3, 2 (2019), 151–160. https://doi.org/10.1080/24751448.2019.1640526

Krzysztof Krejtz, Arzu Çöltekin, Andrew T. Duchowski, and Anna Niedzielska. 2017. Using coeffcient K to distinguish ambient/focal visual attention during cartographic tasks. *Journal of Eye Movement Research* 10, 2 (2017), 1–13. https://doi.org/10.16910/jemr.10.2.3

Krzysztof Krejtz, Andrew Duchowski, Izabela Krejtz, Agnieszka Szarkowska, and Agata Kopacz. 2016. Discerning Ambient/Focal Attention with Coefficient K. *ACM Transactions on Applied Perception* 13, 3 (2016), 1–20. https://doi.org/10.1145/2896452

Jakub Krukar and Ruth Conroy Dalton. 2013. Spatial Predictors of Eye Movement in a Gallery Setting. In *Eye Tracking for Spatial Research, Proceedings of the 1st International Workshop (in conjunction with COSIT 2013)*, Peter Kiefer, Ioannis Giannopoulos, Martin Raubal, and Mary Hegarty (Eds.). Scarborough, UK, 14–19.

Jakub Krukar and Ruth Conroy Dalton. 2020. How the Visitors' Cognitive Engagement Is Driven (but Not Dictated) by the Visibility and Co-visibility of Art Exhibits. *Frontiers in Psychology* 11 (mar 2020). https://doi.org/10.3389/fpsyg.2020.00350

S. F. Kuliga, T. Thrash, R. C. Dalton, and C. Hölscher. 2015. Virtual reality as an empirical research tool - Exploring user experience in a real building and a corresponding virtual model. *Computers, Environment and Urban Systems* 54 (2015), 363–375. https://doi.org/10.1016/j.compenvurbsys.2015.09.006

Yi Lu and Yu Ye. 2019. Can people memorize multilevel building as volumetric map? A study of multilevel atrium building. *Environment and Planning B: Urban Analytics and City Science* 46, 2 (feb 2019), 225–242. https://doi.org/10.1177/2399808317705659

Daniel Müller-Feldmeth, Sarah Schwarzkopf, Simon J. Büchner, Christoph Hölscher, Gregor Kallert, Rul von Stülpnagel, and Lars Konieczny. 2014. Location Dependent Fixation Analysis with Sight Vectors. Locomotion as a Challenge in Mobile Eye Tracking. In *Proceedings of the 2nd International Workshop on Eye Tracking for Spatial Research co-located with the 8th International Conference on Geographic Information Science (GIScience 2014)*, Peter Kiefer, Ioannis Giannopoulos, Martin Raubal, and Antonio Krüger (Eds.). Vienna, Austria, 67–71.

Christina Ohm, Manuel Müller, and Bernd Ludwig. 2017. Evaluating indoor pedestrian navigation interfaces using mobile eye tracking. *Spatial Cognition & Computation* 17, 1-2 (jan 2017), 89–120. https://doi.org/10.1080/13875868.2016.1219913

Christina Ohm, Manuel Müller, Bernd Ludwig, and Stefan Bienk. 2014. Where is the Landmark? Eye tracking studies in large-scale indoor environments. In *2nd International Workshop on Eye Tracking for Spatial Research co-located with the 8th International Conference on Geographic Information Science (GIScience 2014)*, Peter Kiefer, Ioannis Giannopoulos, Martin Raubal, and Antonio Krüger (Eds.). Vienna, Austria, 47–51.

Helmut Schrom-Feiertag, Volker Settgast, and Stefan Seer. 2017. Evaluation of indoor guidance systems using eye tracking in an immersive virtual environment. *Spatial Cognition and Computation* 17, 1-2 (2017), 163–183. https://doi.org/10.1080/13875868.2016.1228654

Stadler and Cornet. 2020. Wayfinding and interaction concepts for an autonomous bus system in Singapore. (2020).

Alasdair Turner, Maria Doxa, David O'Sullivan, and Alan Penn. 2001. From isovists to visibility graphs: a methodology for the analysis of architectural space. *Environment and Planning B: Planning and Design* 28, 1 (2001), 103–121.

Boris M Velichkovsky, Markus Joos, Jens R Helmert, and Sebastian Pannasch. 2005. Two visual systems and their eye movements: Evidence from static and dynamic scene perception. In *Proceedings of the XXVII conference of the Cognitive Science Society*. Lawrence Erlbaum Associates, Mahwah, NJ, 2283–2288.

Flora Wenczel, Lisa Hepperle, and Rul von Stülpnagel. 2017. Gaze behavior during incidental and intentional navigation in an outdoor environment. *Spatial Cognition & Computation* 17, 1-2 (2017), 121–142.

Jan M Wiener, Christoph Hölscher, Simon Büchner, and Lars Konieczny. 2012. Gaze behaviour during space perception and spatial decision making. *Psychological research* 76, 6 (2012), 713–729.

Julie Zook and Sonit Bafna. 2016. The feel of space. In *Take One Building: Interdisciplinary Research Perspectives of the Seattle Central Library*. Taylor & Francis, 95.