GazeNav: Gaze-Based Pedestrian Navigation

Ioannis Giannopoulos
ETH Zurich
Institute of Cartography and Geoinformation
igiannopoulos@ethz.ch

Peter Kiefer
ETH Zurich
Institute of Cartography and Geoinformation
pekiefer@ethz.ch

Martin Raubal
ETH Zurich
Institute of Cartography and Geoinformation
mraubal@ethz.ch

ABSTRACT

Pedestrian navigation systems help us make a series of decisions that lead us to a destination. Most current pedestrian navigation systems communicate using map-based turn-by-turn instructions. This interaction mode suffers from ambiguity, its user's ability to match the instruction with the environment, and it requires a redirection of visual attention from the environment to the screen. In this paper we present GazeNav, a novel gaze-based approach for pedestrian navigation. GazeNav communicates the route to take based on the user's gaze at a decision point. We evaluate GazeNav against the map-based turn-by-turn instructions. Based on an experiment conducted in a virtual environment with 32 participants we found a significantly improved user experience of GazeNav, compared to map-based instructions, and showed the effectiveness of GazeNav as well as evidence for better local spatial learning. We provide a complete comparison of navigation efficiency and effectiveness between the two approaches.

Author Keywords

Pedestrian Navigation; Gaze-Based Interaction; Wayfinding; Eye Tracking

ACM Classification Keywords

H.5.2. Information Interfaces and Presentation: User Interfaces

INTRODUCTION

Humans engage in navigation tasks daily, in familiar as well as unfamiliar environments. Navigation is the combination of wayfinding and locomotion, a coordinated movement in the environment that after a series of correct decisions, finally leads to the targeted destination [20].

While navigating in unfamiliar environments, humans utilize assistance aids to help them make correct decisions. Cartographic paper maps - the classical assistance aids - have nowadays been replaced by digital navigation devices. These applications are mostly limited to map-based turn-by-turn instructions, which have several drawbacks. Most critical, a map-based interaction concept requires the user's attention to be directed to the screen which can cause problems when the visual attention is required in the environment, e.g., to avoid obstacles, other pedestrians, or the surrounding traffic [27]. Auditory navigation systems [9], as an alternative to maps, are not an adequate solution since the noise of an urban environment may be disturbing. Moreover, humans are generally reluctant to wear earplugs while navigating because they feel isolated from the environment [23]. As another drawback of several current pedestrian navigation systems, navigators' activities can be recognized (e.g., tourists can be recognized), and their hands are not free for other tasks other than navigating. These factors have an impact on the User Experience (UX), efficiency, and effectiveness of navigation, as well as on the learning of the environment they navigate in.

In this paper we introduce a novel gaze-based interaction concept for pedestrian navigation that avoids many of the mentioned drawbacks. GazeNav follows the *"what you look at is what you get"* [12] principle, providing navigation information if the street the user is looking at is the one to be followed. GazeNav allows for hands-free navigation, without the need to direct the visual attention away from the environment. The concept is based on eye tracking, a methodology commonly used in HCI [14, 30]. Head mounted eye trackers, such as the one used, have become more and more mobile and compact.

A GazeNav prototype was implemented by processing the gaze data from a mobile eye tracker in real-time, and providing a vibro-tactile signal using a smartphone (as a possible feedback method) to inform the user when she is gazing at the correct street. A user study in the VE with 32 participants was performed to compare GazeNav with the map-based turn-by-turn instructions, demonstrating a significant increase in user experience and spatial learning. The GazeNav approach can be effectively used without the need for further information, thus, being able to navigate hands-free without having to interact with a device, but only with the environment.

The contribution of this work is manifold, providing novel insights into navigation assistance systems by giving directions on how gaze can be utilized for:

- dissolving navigation ambiguities,
- hands-free interaction,
- increased usability,
- improved local spatial learning and,
- natural interaction with the environment.

MobileHCI '15, August 25 – 28, 2015, Copenhagen, Denmark
 ISBN 978-1-4503-3652-9/15/08...$15.00
DOI: http://dx.doi.org/10.1145/2785830.2785873

The rest of this paper is structured as follows: we continue with related work and then introduce the GazeNav concept followed by the implementation description. Next, the experiment is introduced followed by the results section. Finally, we close with a discussion and objectives for future work.

RELATED WORK

Gaze-Based Interaction

Eye trackers measure the visual attention on a stimulus [3]. The basic eye tracking measurement is a (x,y) coordinate pair in the coordinate system of the eye tracker (called gaze). These gaze points are typically aggregated to fixations, which are the moments when perception is assumed to take place [24]. The transition between two fixations is called a saccade.

By processing the eye tracking data in real-time, interfaces that react to a user's visual attention become possible [36]. Some approaches have implemented gaze-based interaction for desktop applications with the eye tracker mounted below the screen. Head-mounted eye tracking hardware, in contrast, has facilitated mobile gaze-based interaction [6], where the interaction may also take place with objects in a 3D environment, such as exhibits in a museum [33]. In this paper, we introduce gaze-based interaction with street junctions in an urban environment.

Gaze-based interaction can be used to replace other interaction methods, or as part of multimodal interaction [29]. It has been argued that supplementing gaze with other interaction modalities prevents the Midas touch problem [12, 13], which may occur if the user's gaze triggers an unwanted interaction.

There are, in principle, two ways of designing gaze-based interaction: with explicit or implicit interaction. Explicit interaction occurs when the user can trigger an interaction by intentionally fixating at a certain position or by performing a gaze gesture [10]. We call an interaction implicit if it records the user's gaze during regular interaction and uses this information to adapt to the user's needs at some later time (e.g. [6]). The GazeNav interaction design proposed in this paper uses explicit interaction with a haptic vibration feedback. Kangas et al. [15] found that the maximum possible delay between a gaze and a vibration, to ensure the user can still identify the two with each other, is around 250 ms.

Pedestrian Navigation Assistance

In the past, mobile devices have been used to provide pedestrian navigation assistance, mostly using a map-based turn-by-turn instructions approach. This kind of navigation forces the user to spend attention on the map instead of letting her interact with the environment (e.g., to avoid obstacles or enjoy the view). In recent years, researchers have tried to overcome these problems by introducing novel interaction methods [17, 26, 27, 37, 38]. There have been approaches to support navigation through play [1], using augmented reality [31], through auditory interfaces [9], or even through music [7]. These approaches perform well in enriching the pedestrian guidance with a stimulating sensory experience. In many pedestrian navigation scenarios, however, the system should be as non-distractive as possible. This is exactly the goal of GazeNav.

One approach that overcomes many of these problems is a vibro-tactile waist belt [35, 22]. This interaction approach informs the user about the direction she has to take by tactile feedback. Similarly, a vibrating smartphone as tactile feedback is used by PocketNavigator [23]. While these systems are privacy-preserving and non-distractive, they suffer from a rather high ambiguity and may require a long learning phase, since the user needs to map the characteristic of the vibration (e.g., its position, intensity, or rhythm) to an executable wayfinding decision. Ambiguity can lead to incorrect decisions and increase uncertainty, which in turn leads to insecure navigators not being able to validate their decision. Altogether, these problems can lead to a generally bad UX.

Utilizing human gaze can provide easy, fast and natural ways of interaction [32] leading to gaze-based pedestrian navigation systems that can avoid the problem of ambiguity and the problem of attention switches, minimizing the interaction with the device and improving the user experience.

In this work we introduce such a gaze-based pedestrian navigation system called GazeNav. We utilize the user's gaze to inform her if she has to take a turn at the street she is looking at. Currently we support a vibro-tactile feedback using a smartphone, similar to the approaches described above. In our approach, however, ambiguity is removed through the direct mapping of gaze to feedback: "what you look at is what you get" [12].

Spatial Knowledge

Spatial knowledge is assumed to be represented in a cognitive map [5], a mental representation that corresponds to people's perception of the real world, although other metaphors, such as cognitive collages and spatial mental models, have also been proposed [34]. Different types of spatial knowledge can be distinguished, and are assumed to be acquired in this order: knowledge about landmarks, followed by knowledge about routes connecting these landmarks, finally leading to configuration knowledge about the spatial relations between landmarks independent from routes [19]. Evidence was also shown though that spatial knowledge acquisition may not necessarily take place in such fixed order [11].

Pedestrian navigation systems are specifically helpful in areas where we have not (yet) acquired spatial knowledge. Although this is very convenient, it is a well-known problem that "[o]ver-reliance on the automated system may cause users to be 'mindless' of the environment and not develop wayfinding and orientation skills nor acquire the spatial knowledge that maybe required when automation fails." [21, p. 238]. Researchers have thus highlighted the need for designing pedestrian navigation systems in a way that they support spatial learning [25].

GazeNav intends to assist its users in a non-distracting way, allowing them to keep their visual attention to the environment without interruption. In our study we will test with a scene recognition test whether this interaction concept supports spatial learning of landmark knowledge.

GAZENAV

Visual Attention Switches and Instruction Ambiguity

The GazeNav concept approaches the problem of attention switches between the real world and the mobile device used for navigation as well as the problem of navigation instruction ambiguity. Consider the following example scenario:

Bob has just arrived in X-town, a city he has never visited before. He starts a navigation application on his mobile device and types in the address of the hotel he has booked. Bob has a hard time figuring out how to hold the mobile device in his hands, in order to be able to look at the navigation instructions. He is carrying two big suitcases with him.

The example above describes a typical scenario for current standard mobile pedestrian navigation systems. Users are not effectively supported in information requests that relate to the navigation instructions. Visual attention switches to the mobile device become necessary in order to make a decision, thus the necessity of holding the device in the hand occurs. Next to this necessity, the user has to be able to clearly understand and process the instructions in order to make a correct decision:

Bob is getting closer to a decision point. He interrupts locomotion and starts using the device in order to read the next navigation instruction. The instruction is a left turn. Bob is struggling to make a decision since there are several possible left turns very close to each other. Bob tries to find hints on the map and in the environment that could possibly be matched in order to dissolve the ambiguity.

Depending on the goal we have while navigating, an incorrect decision can be critical. Before deciding to take a direction, humans try to validate the instructions based on several environmental factors, such as landmarks and the geometry of the streets.

GazeNav Interaction Concept

In our interaction concept we enable a gaze-based interaction for pedestrian navigation. We utilize the gaze in order to inform the navigator when she is gazing at a street where a turn is necessary. Consider the following application scenario:

Bob types the desired destination into the GazeNav app. He puts the device in his pocket, puts on his eye tracking-enabled glasses, picks up his two suitcases, and starts walking. At the first street junction he gazes at a street to his left, when the phone in his pocket starts vibrating. Bob immediately knows he has to follow that specific street, and turns left without further hesitation.

With GazeNav we try to overcome the problems of the existing navigation technologies by incorporating the human gaze in the interaction dialog while navigating. Our hypotheses are as follows:

H1. (Effectiveness) As GazeNav prevents disambiguation errors, users of the system will be able to reach their destination without the need for further assistance aids.

H2. (Efficiency) Navigation efficiency, in terms of navigation duration, with GazeNav is at least as good as with map-based turn-by-turn instructions.

H3. (Spatial learning) Since the visual attention always stays on the environment, users of GazeNav will acquire better local spatial knowledge about the environment than users of map-based turn-by-turn instructions.

H4. (User Experience) The User Experience (UX) of GazeNav will be better than that of map-based turn-by-turn instructions.

IMPLEMENTATION

We implemented a prototype of the GazeNav interaction concept and used it for the experiment described in the next section.

Hardware

The used hardware consisted of the Ergoneers Dikablis[1] mobile eye tracker with a gaze capture rate of 25Hz. The gaze data were transmitted to a laptop via a coaxial cable. The laptop forwarded the data to a Samsung Galaxy Nexus smartphone where the prototype application was running. The connection between laptop and smartphone was established over Wi-Fi in a closed network.

The experiment took place in a virtual environment projected to a wall using an EPSON projector with a resolution of 1920 x 1200 pixels. A Logitech 3D precision pro joystick was utilized with which participants could navigate through the virtual environment. For the control condition, we used the same set-up plus a 19" screen to display the map-based turn-by-turn instructions. The size of the screen was selected in order to optimize the tracking process of the visual attention.

Software

Calibration and Recording (Laptop)

The software modules that come together with the Dikablis mobile eye tracking system were used for the calibration, as well as to control the recording sessions. A real-time module that provides visual marker detection was used, returning gaze points in a marker coordinate system with respect to markers attached in the virtual environment. These coordinates were transmitted over UDP to the smartphone.

Virtual Environment

The ESRI CityEngine[2] was utilized to design a virtual environment. Using this software we could generate a random urban environment and manually incorporate the desired path for the navigation experiment.

In a second step, the virtual environment was imported to the Unity[3] game engine, which made the generated city navigable in first-person view with a joystick (see Figure 1).

In the last step, we attached markers to the buildings along the paths the users could potentially take during the study. These

[1]http://www.ergoneers.com/

[2]http://www.esri.com/software/cityengine

[3]http://unity3d.com/

Figure 1. A scene of the virtual environment used for the experiment. It was created using the ESRI CityEngine and Unity Game Engine.

Figure 2. Turn-by-turn instructions visualized on a map were used in the second condition of the experiment.

markers were used for image recognition purposes. Using markers made it easier and faster to test the prototype implementation. In a real-world scenario, image detection would have to be performed on images of buildings.

GazeNav Application (Smartphone)
The GazeNav app was implemented with Android (Android 4.2.1) and is needed for the activation of the vibro-tactile feedback on the smartphone.

The marker dependent coordinates received over UDP were used to check if the user was looking at the street she was supposed to take. In that case, the phone started vibrating. Vibration was used as a signaling method, since it is non-distracting, privacy-preserving, and has been used successfully in other approaches [23, 35]. Our interaction concept is not bound to vibration as a signaling method and the signaling was not the main scope of our research.

EXPERIMENT
We used the implementation described in the previous section and performed a user study to compare two pedestrian navigation systems, our gaze-based approach GazeNav and a map-based turn-by-turn instructions approach, that served as the baseline since it is one of the mainly used navigation types. The hypotheses tested are the ones introduced in the section GazeNav.

The map-based turn-by-turn instructions approach is the typical one used for pedestrian navigation on mobile devices (e.g., GoogleMaps or OpenStreetMaps pedestrian navigation). In this condition, participants had to navigate using the instructions presented on the screen. The instructions were in the form of direction arrows presented together with an allocentric map, using a purple dot denoting the current location of the user in the urban environment and an egocentric heading. The path to follow was highlighted, and the instruction was displayed on the top left of the map (see Figure 2). We implemented a prototype for this condition, but in order to save implementation time, for the updating of the map and the location of the user we used the "Wizard of Oz" experiment methodology [16], having the experimenter update the map, based on the location of the user in the VE.

The participants' visual attention on the instruction screen was tracked during the experiment. Rotation of the map was not necessary and was not asked for by any of the participants, since the displayed map had an egocentric heading and directional cues were present.

Setup
Participants were placed in front of a height-adjustable table with a gaming joystick in the center and a smartphone on the side (see Figure 3, left). The table was positioned at a distance of 3 meters in front of a projection wall (see Figure 3, right). For the test condition where the participants had to navigate using turn-by-turn instructions, also a 19" screen was positioned in front of them (~ 15 deg. angle).

Right before the experiment, the mobile eye tracker was mounted on the participant's head and calibrated to the distance of the projection screen.

Design
A between subjects design was employed to compare the two pedestrian navigation systems. Each participant from both groups had to navigate along the same route in the same virtual environment. The only difference between the two groups was the assistance system used.

We designed a route with several decision points, having a different number of connections (see Figure 4) with the intention to investigate the effect of decision points with varying structure and complexity. Since the tested route covered different levels of complexity concerning the decision points as well as navigation directions towards all cardinal directions, there was no need to perform the experiment using a second route.

Procedure
Each experiment, in both conditions, was composed of five steps. At first, participants had to provide their demographic information, their experience level with mobile navigation systems and gaming joysticks as well as to fill in a questionnaire for the self-estimation of their spatial abilities [8].

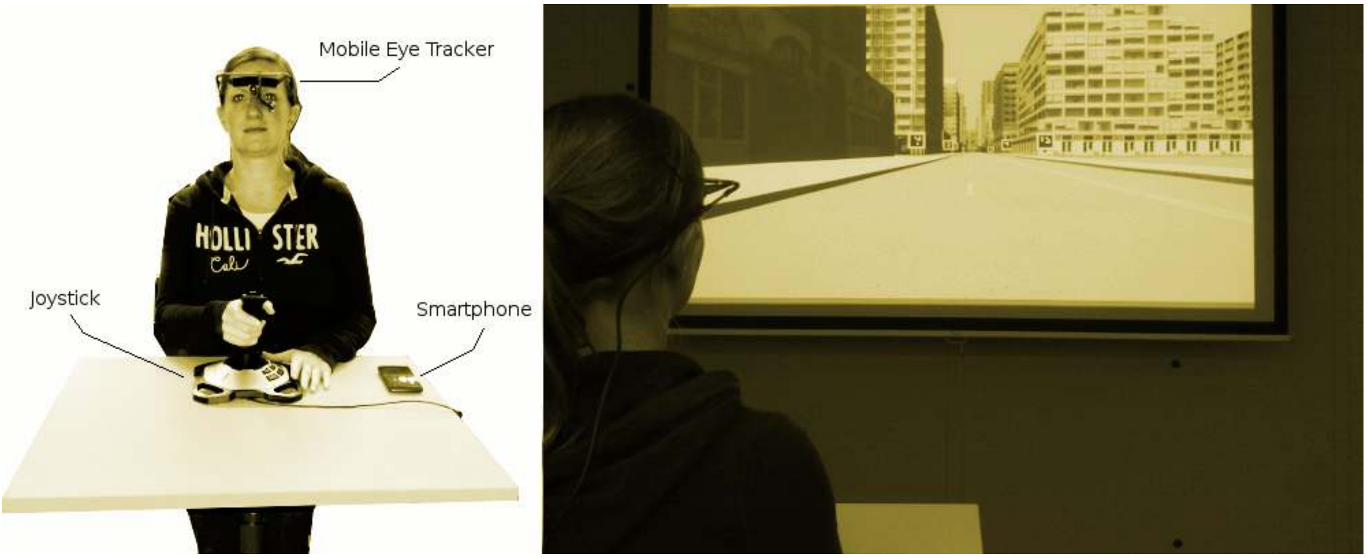

Figure 3. Experimental setup. A participant, equipped with mobile eye tracker (left), and setup in the virtual environment (right).

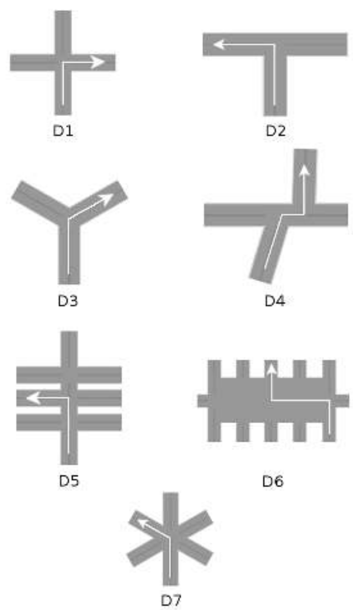

Figure 4. This figure illustrates the decision points, D1 to D7, where the participant had to take a turn.

In the second step, participants could try out the navigation assistance system they were assigned to as well as get used to the gaming joystick by navigating in a virtual environment, different to the one used in the next step.

The third step was the actual trial. Each trial started in the same virtual environment, using a first-person view. The participants were told to use the gaming joystick and navigate through the environment, trying to find the target destination as fast as possible.

In the fourth step of the experiment, immediately after the trial, participants were given a set of 17 printouts (11 correct) with egocentric images of street intersections from the virtual environment. Their task was to choose the printouts with the intersections they thought to remember having crossed during navigation and also sort them in the order they occurred in the environment. In the final step, the participants had to evaluate the user experience of the tested system by filling in a standardized questionnaire ([18], UEQ).

Tracked Data

In both conditions we tracked the completion time (i.e., the time participants needed to reach the target destination), the time after the last decision until the new decision was made (movement speed was constant), the number of times a participant stopped locomotion (e.g., to look on the map in the Turn-by-Turn condition) as well as the number of attention switches (e.g., from the environment to the map in the Turn-by-Turn condition). The time was measured through our implementation whenever the participant reached predefined points on the route.

Pilot Study

A pilot study with four participants was conducted in order to test the stability of the virtual environment and the navigation systems for both conditions as well as to optimize the experimental procedure. The insights retrieved from the pilot study helped us determine the trigger area for the GazeNav system (i.e., the bounding box around the streets used as a gaze-based trigger region) as well as the duration of the vibration response.

In order to determine the trigger area we let the pilot users navigate through the virtual environment. The trigger areas varied at every street the participants had to follow. Finally, we were able to identify one trigger area that worked well for our experiment. We used a trigger area that spans over a street and is 4 meters high. We do not claim that the selected trigger area is the optimal one to use in every environment. More studies need to be conducted in order to develop an algorithm for the determination of an optimal trigger area, but this is outside the scope of this work.

For the determination of the response vibration duration, we tried several thresholds. We tested thresholds of 100, 200, 300, 400, and 500 milliseconds as well as a continuous vibration (i.e., vibration as long as the user is gazing at the right street). Vibration responses using a time threshold caused confusion, since the human eye is able to scan an area very fast, making it difficult to match the vibration response with the actual gaze point. The continuous vibration approach worked well, since the users were able to easily match the response with the correct street. Based on these findings we decided to use a continuous vibration response. One common user feedback concerning the continuous vibration response was that it also served as a validation, allowing the users to verify several times that the street they were heading to was the correct one.

Participants

In total 32 participants were recruited for the experiment in exchange for a small honorarium, 16 participants for each condition. The participants had different cultural (e.g., Europe, US) and professional backgrounds (e.g., physicists, artists, computer scientists). The participants of the GazeNav condition (11 male, 5 female) had a mean age of 32.81 years (SD = 6.18). The participants of the Turn-by-Turn condition (8 male, 8 female) had a mean age of 31.13 years (SD = 8.09). The participants of the GazeNav condition also had to wear a head-mounted mobile eye tracking device that had to be calibrated. The time necessary to wear and calibrate the device was not considered in the experiment.

Mobile Navigation Systems Expertise

Participants of the GazeNav condition rated their experience using mobile navigation systems with a mean of 5.68 (SD = 1.31), with seven noting the highest experience on a 7-point Likert scale. Participants of the Turn-by-Turn condition rated their experience with a mean of 4.75 (SD = 1.69). A Mann-Whitney U test did not reveal a significant difference between the expertise of the two groups (p = .164, Z = -1.391).

Gaming Joystick Experience

Participants in both conditions were also asked to rate their experience with gaming joysticks. On a 7-point Likert scale, participants of the GazeNav condition rated their experience with a mean of 2.62 (SD = 2.33) and the participants of the Turn-by-Turn condition with a mean of 2.56 (SD = 2.06). A Mann-Whitney U test did not reveal a significant difference between the expertise of the two groups (p = .787, Z = -.271).

Spatial Abilities

The participants were asked to fill in the "Santa Barbara Sense of Direction Scale" questionnaire [8], which is a self-estimation of spatial abilities. It is a measure for environmental spatial ability that correlates with objective measures of performance in several environmental spatial cognition tasks [8]. Participants had to rate their abilities using a 7-point Likert scale, answering questions such as *"I have trouble understanding directions"*. A Mann-Whitney U test showed that the self-estimated spatial abilities of the participants between the two tested groups were not significantly different (p = .157, Z = -1.415). The participants in the GazeNav condition had a mean of 5.32 (SD = 1.12), and the participants of the Turn-by-Turn condition had a mean of 4.92 (SD = 1.01). A score close to 7 indicates high spatial abilities.

RESULTS

We analyzed the data collected during the experiment in order to investigate the hypotheses stated in the GazeNav section.

Navigation Performance

One measurement for navigation performance is the time necessary in order to reach the destination (H2). The total time was not significantly different between the groups, but there were some significant differences when looking at the individual segments of the route (see Table 1). These detailed analyses were necessary in order to validate the overall result and to exclude the possibility that the result was caused only by a single decision point. Participants using GazeNav navigated significantly faster on the first segment, i.e., between the starting point and the first decision point (D1) where the participant had to take her first turn, as well as on the following segment between decision points D1 and D2. Participants navigating using the turn-by-turn instructions were only significantly faster at one segment: between decision points D5 and D6.

Concerning navigation effectiveness (H1), all participants in both conditions reached the target destination, but participants using the turn-by-turn instructions often interrupted locomotion in order to read the navigation instructions. On average, they stopped 15.5 times (min = 0, max = 36, SD = 14.56). Participants using GazeNav did not interrupt their locomotion at all.

Visual Attention Switches

The participants in the condition using the GazeNav navigation system spent 100% of their visual attention on the virtual environment. In contrast, the participants of the turn-by-turn condition had to look at the instructions in order to make decisions, leading to several attention switches from the virtual environment to the display and vice versa. The mean number of attention switches was 42.06 (min = 19, max = 83, SD = 18.36), and the mean total time spent looking at the instructions was 14.5 seconds (min = 2, max = 40, SD = 11.07).

Spatial Learning

The analysis of the spatial learning task (refer to H3) revealed a significant difference between the two tested conditions concerning correct selections. Participants in the GazeNav condition had a mean of 7.31 correct selections (min = 4, max = 10, SD = 1.85). Participants of the turn-by-turn condition had a mean of 5.81 correct selections (min = 4, max = 10, SD = 2.073). This difference was analyzed using a Mann-Whitney U test, showing a significant difference between the two conditions (p < .05, Z = -2.155). Concerning the number of incorrect selections, participants of the GazeNav condition had a mean of 1.13 incorrect selections (min = 0, max = 4, SD = 1.25) and participants of the turn-by-turn condition had

		D1	D2	D3	D4	D5	D6	D7	Total Time
1	**Mean**	44.6	79.3	59.3	30.9	72.2	75.9	16.7	379.3
2		50.1	83.6	58.5	30.0	66.3	48.4	17.2	354.1
1	**SD**	11.4	3.4	8.6	11.9	9.8	29.7	3.3	0.7
2		8.3	4.1	4.2	6.6	4.9	6.7	4.1	0.4
Mann-Whitney U		*Z=-2.751*	*Z=-3.090*	*Z=-.754*	*Z=-.942*	*Z=-1.545*	*Z=-2.224*	*Z=-.905*	*Z=-.696*
		p <0.01	p <0.01	p = .468	p =.361	p = .128	p <.05	p = .381	p = .094

Table 1. The table depicts the performance (seconds) of the participants in the two conditions (1: GazeNav, 2: Turn-by-Turn). The values in the cells represent the descriptive and inferential statistics on the duration from one decision point to the next.

a mean of 1.06 incorrect selections (min = 0, max = 3, SD = .85). This difference was not significant (p = .76, Z = -.301).

In order to analyze the impact of the selected order, we used the edit distance algorithm of Levenshtein. The selections of the participants in the GazeNav condition needed on average 7.69 edits in order to match their selection with the correct one (min = 4, max = 10, SD = 1.58). The selections of the participants in the turn-by-turn condition needed on average 7.5 edits (min = 6, max = 10, SD = 1.15). There was no significant difference between the two groups (p = .59, Z = -.530).

User Experience

Concerning the user experience (H4), the GazeNav approach performs better than the turn-by-turn approach for all scales (see Figure 5). A Mann-Whitney U statistical test also showed that the differences are statistically significant for Attractiveness, for the pragmatic scales Perspicuity and Efficiency, as well as for the hedonic scales Stimulation and Novelty (see Table 2).

Mann-Whitney U		
Attractiveness	p <.001	Z = -2.951
Perspicuity	p <.01	Z = -2.608
Efficiency	p <.001	Z = -3.606
Dependability	p = .518	Z = -.646
Stimulation	p <.01	Z = -3.329
Novelty	p <.001	Z = -4.071

Table 2. Inferential statistics for the UX comparison between the two tested conditions.

In order to investigate if the user experience of GazeNav is sufficiently high to fulfill the general expectation of users, we compared our results with a benchmark dataset [28]. These user expectations are strongly influenced through the user interaction with other systems.

GazeNav performed excellent in the benchmark (see Figure 6), having 5 scales of the user experience in the range of the 10% of best results and one scale, Dependability, above average, with only 25% of the results in the benchmark being better.

DISCUSSION

All Hypotheses stated at the GazeNav section could be confirmed. Our experiment clearly revealed the benefits of the GazeNav approach against the map-based turn-by-turn instructions and at the same time highlighted the main disadvantages of navigation aids that require visual attention and a minimum of spatial abilities, giving directions on possible solutions.

Concerning the effectiveness (H1) of the two navigation systems, both fulfilled their purpose, guiding their users to the target destination. In other words, our GazeNav concept was not worse than the turn-by-turn concept in terms of effectiveness. Moreover, one of the benefits of GazeNav is that the participants did not have to interrupt their locomotion as opposed to the baseline approach.

The total navigation duration (H2) between the two conditions did not reveal significant differences. The results demonstrated that GazeNav performed significantly better at two decision points, and significantly worse at one decision point. The better performance of turn-by-turn instructions at D6 (the square, refer to Figure 4) was due to image recognition problems in the GazeNav condition. Depending on the angle the user was trying to gaze at a street, the visual markers that were adjusted to the buildings could not always be detected by the marker detection software (due to directional light). This software problem forced the GazeNav users to walk around the square, trying to find the correct street to follow. This is also the reason why GazeNav did not perform better than the turn-by-turn condition, although the participants of the baseline condition interrupted their locomotion several times.

The users of the GazeNav approach did not have to interact with anything else except for the environment they were navigating in. The Turn-by-Turn instructions, in contrast, forced the users to switch their attention several times to the navigation device. Moreover, while looking at the instructions, many participants interrupted their locomotion. Thus, our results underline that GazeNav is non-distracting, which was the main design goal of our approach.

The scene recognition task (H3) revealed significantly better results for the participants using GazeNav concerning the number of correctly recognized scenes, suggesting better local spatial learning. The ordering task though, did not reveal any significant differences.

The participants were excited about using GazeNav, which is also reflected in the results of the UEQ questionnaire. Navigation using GazeNav clearly enhances the user experience. According to the results of the UEQ questionnaire, GazeNav outperformed the map-based turn-by-turn instructions and also performed excellent w.r.t. the used benchmark. Concerning the user feeling of being in control of the interaction, GazeNav performed better, but not significantly.

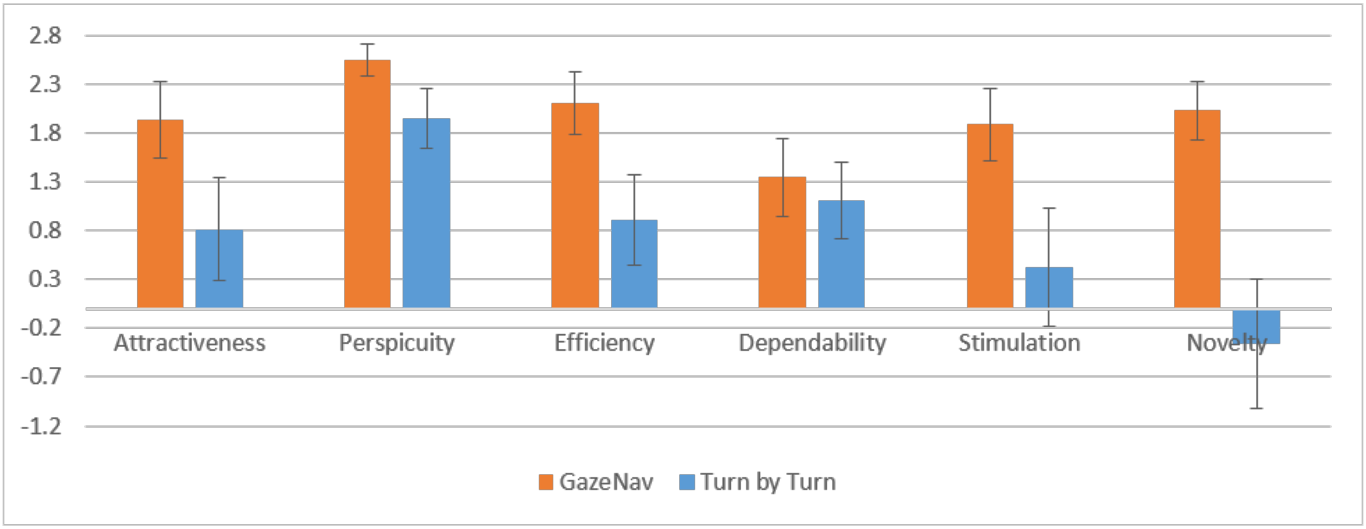

Figure 5. User experience evaluation.

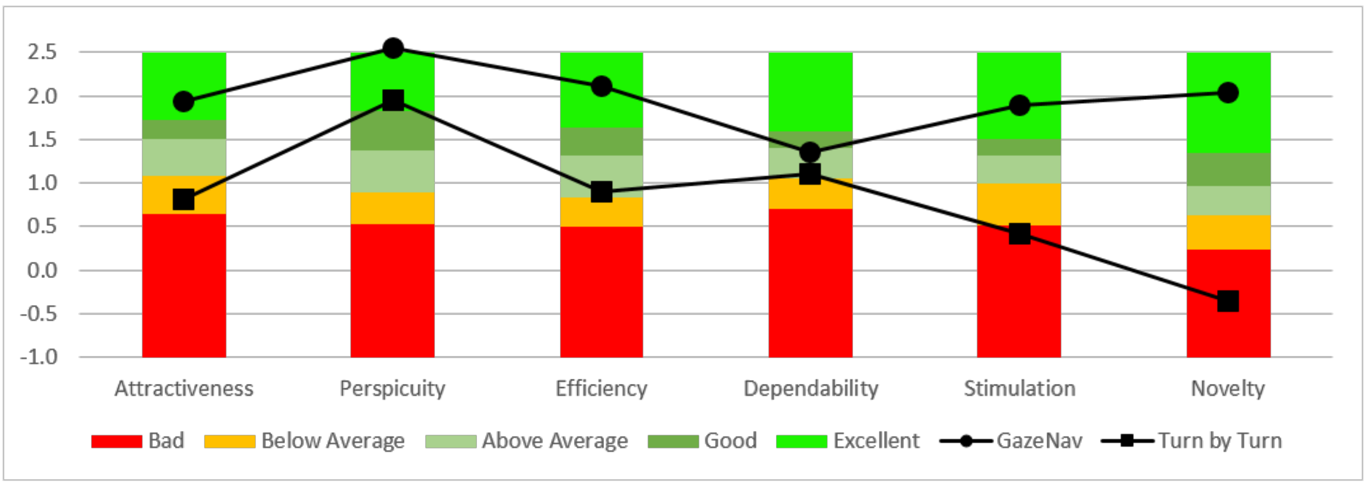

Figure 6. User experience benchmark.

A very important situation that has to be considered by a system like GazeNav is the case when a correct turn is overseen and the user is walking in the wrong direction. The system should inform the user that she is walking towards a wrong direction. This can be easily achieved by utilizing the existing feedback modality, vibration. For instance by providing an easy to perceive vibro-tactile pattern feedback.

When using a map for navigation, one advantage for the user is that she can plan ahead. This advantage can still be valid for GazeNav. The user can use her smartphone to type in the destination using a routing service and before putting the smartphone back into her pocket, take a look at the route, thus being also able to plan ahead. GazeNav would serve as a memory augmentation, helping the user remember and validate the right turns.

Limitations

The external validity of lab studies is always an issue when performing experiments. Studies in virtual environments cannot replicate all contextual factors of outdoor environments (e.g., weather conditions) and we are aware of the ongoing debate concerning the external validity of lab studies [2, 4]. One of the main advantages of a lab setting is the control over the factors that are relevant, whereas the main benefit of field studies is the ecological validity [2].

For the evaluation of the GazeNav concept, we chose a controlled experimental setup in a virtual environment in order to (1) *control for technological limitations* as well as (2) *to vary the complexity of the environment*.

(1) Our focus is on the interaction concept, not on improving mobile eye tracking (ET) technology. In a real-world study we may find out that participants perform better in cloudy weather. This would be an evaluation of the mobile ET base technology, but not of GazeNav. Hardware limitations are ongoing research in other domains and not within the scope of this paper.

(2) The virtual environment enabled us to systematically vary the complexity of decision points (see Figure 4). We argue that the validity of the results can be better controlled for if factors such as weather conditions are excluded.

Another limitation of navigation studies in the lab is the unnatural movement, using a joystick. Although the joystick experience is different to the natural movement, it is a very commonly used input device to simulate movement and the participants of both experiment groups were equally experienced joystick users. Although we argue that one of the benefits of GazeNav relies in the hands-free interaction, we had no other possibility for movement. Using a joystick in this experiment actually shows that the users' hands can be utilized for other tasks, since in a setup that would not require simulated movement, the hands would be free.

Since both conditions were tested in the same controlled environment, the limitations apply to both conditions. Nevertheless, a comparative study in a real environment would reveal the full potential of GazeNav and is our focus for future research.

CONCLUSIONS AND FUTURE WORK

We have presented GazeNav, a novel interaction concept for pedestrian navigation systems that allows the user to interact with the environment by using gaze. Following the "what you look at is what you get" [12] principle, a user approaching a decision point is provided with feedback about whether the street she is looking at is the one to follow. We argued that GazeNav has advantages over map-based turn-by-turn instructions: it is operated hands-free and allows the user to keep the visual attention to the environment. Moreover, the direct feedback to gaze makes navigation decisions unambiguous, compared to maps which require a time consuming and error-prone cognitive matching process between instructions and the real world.

In a study with 32 participants in a virtual environment we found that it is possible to effectively navigate with GazeNav (all participants reached their goal). We compared GazeNav with map-based turn-by-turn instructions and found significantly higher user experience scores for GazeNav. The user experience of our system was also excellent when compared to a benchmark dataset [28]. Participants with GazeNav were able to acquire significantly better local spatial knowledge, which could be explained by them being able to keep their visual attention to the environment. Concerning the efficiency, there was no significant difference between the two interaction concepts, strengthening our hypothesis, that navigation with GazeNav where no further aids are necessary, will not have an efficiency loss.

In future work, we will implement the system for the real world. Image processing techniques will be used to localize the image of the front-facing eye tracking camera, thus enabling a mapping of gaze to the trigger areas. Automatically attaching trigger areas to streets at crossroads is another issue to optimize in future work (as this placement was not our focus here, we determined the position of trigger areas through a pilot study). The only hardware that would be necessary for a user to carry, would be a mobile eye tracking device and a smartphone to serve for vibro-tactile feedback.

It will further be interesting to compare GazeNav with other interaction methods, such as vibrating belts [35] or head-mounted displays (e.g., Google Glass). We hypothesize that, due to the direct coupling of gaze direction and vibration, GazeNav is less ambiguous than a vibrating belt, and provides a more seamless experience than Google Glass. Hybrids are also in the focus of this research, combining GazeNav with approaches such as vibrating belts will have the advantage of dissolving the ambiguities (through gaze) and at the same time minimizing the search space for intersections (through e.g., vibrating belts). Determining the optimal feedback method, other than vibration, is also in focus for future work.

GazeNav was presented in the context of pedestrian navigation. Although not evaluated, it might be possible to generalize GazeNav and apply it to other types of navigation where decisions have to be made faster, e.g., car or bicycle navigation. For instance, GazeNav could be integrated into the existing map-based car navigation systems in order to help the driver validate her decision faster. This generalization of the presented concept will also be a direction for future work, trying to identify the challenges posed by these types of navigation.

REFERENCES

1. Marek Bell, Stuart Reeves, Barry Brown, Scott Sherwood, Donny MacMillan, John Ferguson, and Matthew Chalmers. 2009. EyeSpy: Supporting Navigation Through Play. In *CHI*. ACM, 123–132.
2. Ioannis Delikostidis, Holger Fritze, Thore Fechner, and Christian Kray. 2014. Bridging the Gap Between Field- and Lab-Based User Studies for Location-Based Services. In *Progress in Location-Based Services*. Springer International Publishing, Chapter 18, 257–271.
3. Andrew T Duchowski. 2007. *Eye Tracking Methodology: Theory and Practice* (2nd ed.). Springer.
4. Henry Been-Lirn Duh, Gerald CB Tan, and Vivian Hsueh-hua Chen. 2006. Usability evaluation for mobile device: a comparison of laboratory and field tests. In *MobileHCI*. ACM, 181–186.
5. Tommy Garling, Anders Book, and Erik Lindberg. 1984. Cognitive Mapping of Large-Scale Environments: The Interrelationship of Action Plans, Acquisition, and Orientation. *Env. and Behavior* 16, 1 (1984), 3–34.
6. Ioannis Giannopoulos, Peter Kiefer, and Martin Raubal. 2012. GeoGazemarks: Providing gaze history for the orientation on small display maps. In *ICMI*. ACM, 165–172.
7. Adrian Hazzard, Steve Benford, and Gary Burnett. 2014. Walk this Way : Musically Guided Walking Experiences. In *CHI*. ACM, 605–614.
8. Mary Hegarty, Anthony E Richardson, Daniel R Montello, Kristin Lovelace, and Ilavanil Subbiah. 2002.

Development of a self-report measure of environmental spatial ability. *Intelligence* 30, 5 (2002), 425–447.

9. Simon Holland, David R Morse, and Henrik Gedenryd. 2002. AudioGPS: Spatial audio navigation with a minimal attention interface. *Personal and Ubiquitous computing* 6, 4 (2002), 253–259.
10. Aulikki Hyrskykari, Howell Istance, and Stephen Vickers. 2012. Gaze gestures or dwell-based interaction?. In *ETRA*. ACM, 229–232.
11. Toru Ishikawa and Daniel R Montello. 2006. Spatial knowledge acquisition from direct experience in the environment: Individual differences in the development of metric knowledge and the integration of separately learned places. *Cog. psychology* 52, 2 (2006), 93–129.
12. Robert J K Jacob. 1990. What you look at is what you get: eye movement-based interaction techniques. In *CHI*. ACM, 11–18.
13. Robert J K Jacob and Keith S Karn. 2003. Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises. In *The Minds Eye: Cognitive and Applied Aspects of Eye Movement Research*, Hyona, Rachach, and Deubel (Eds.). Elsevier, Oxford, England, 573–605.
14. Jari Kangas, Deepak Akkil, Jussi Rantala, Poika Isokoski, Päivi Majaranta, and Roope Raisamo. 2014a. Gaze Gestures and Haptic Feedback in Mobile Devices. In *CHI*. ACM, 435–438.
15. Jari Kangas, Jussi Rantala, Päivi Majaranta, Poika Isokoski, and Roope Raisamo. 2014b. Haptic Feedback to Gaze Events. In *ETRA*. ACM, 11–18.
16. J F Kelley. 1983. An Empirical Methodology for Writing User-friendly Natural Language Computer Applications. In *CHI*. ACM, 193–196.
17. Frederic Kerber, Antonio Krüger, and Markus Löchtefeld. Investigating the Effectiveness of Peephole Interaction for Smartwatches in a Map Navigation Task. In *MobileHCI*. ACM, 291–294.
18. Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and evaluation of a user experience questionnaire. *HCI and Usability for Education and Work* 4 (2008), 63–76.
19. Joseph P Magliano, Robert Cohen, Gary L Allen, and James R Rodrigue. 1995. The impact of a wayfinder's goal on learning a new environment: Different types of spatial knowledge as goals. *Environmental Psychology* 15, 1 (1995), 65–75.
20. Daniel R Montello. 2005. Navigation. In *Cambridge handbook of visuospatial thinking.* 257–294.
21. Avi Parush, Shir Ahuvia, and Ido Erev. 2007. Degradation in spatial knowledge acquisition when using automatic navigation systems. In *Spatial information theory*. 238–254.
22. Martin Pielot, Niels Henze, and Susanne Boll. 2009. Supporting map-based wayfinding with tactile cues. In *MobileHCI*. ACM.
23. Martin Pielot, Benjamin Poppinga, and Susanne Boll. 2012. PocketNavigator : Studying Tactile Navigation Systems. In *CHI*. ACM, 3131–3139.
24. K Rayner. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin* 124, 3 (1998), 372–422.
25. Kai-Florian Richter, Drew Dara-Abrams, and Martin Raubal. 2010. Navigating and learning with location based services: A user-centric design. In *LBS and Telecartography*. 261–276.
26. Enrico Rukzio, Michael Müller, and Robert Hardy. 2009. Design, implementation and evaluation of a novel public display for pedestrian navigation. In *CHI*. ACM, 113–122.
27. Sonja Rümelin, Enrico Rukzio, and Robert Hardy. 2011. NaviRadar: A Novel Tactile Information Display for Pedestrian Navigation. In *UIST*. ACM, 293–302.
28. M Schrepp, S Olschner, and U Schubert. 2013. User Experience Questionnaire Benchmark Praxiserfahrungen zum Einsatz im Business-Umfeld. *Usability Professionals* 13 (2013).
29. Sophie Stellmach and Raimund Dachselt. 2012a. Investigating gaze-supported multimodal pan and zoom. In *ETRA*. ACM, 29–36.
30. Sophie Stellmach and Raimund Dachselt. 2012b. Look & touch: gaze-supported target acquisition. In *CHI*. ACM, 2981–2990.
31. Yuichiro Takeuchi and Ken Perlin. 2012. ClayVision: the (elastic) image of the city. In *CHI*. 2411–2420.
32. Vildan Tanriverdi and Robert J K Jacob. 2000. Interacting with Eye Movements in Virtual Environments. In *CHI*. ACM, 265–272.
33. Takumi Toyama, Thomas Kieninger, Faisal Shafait, and Andreas Dengel. 2012. Gaze guided object recognition using a head-mounted eye tracker. In *ETRA*. ACM, 91–98.
34. Barbara Tversky. 1993. Cognitive maps, cognitive collages, and spatial mental models. In *Spatial Information Theory A Theoretical Basis for GIS*. Vol. 716. Springer, Chapter 2, 14–24.
35. Jan BF Van Erp, Hendrik AHC Van Veen, Chris Jansen, and Trevor Dobbins. 2005. Waypoint navigation with a vibrotactile waist belt. *ACM Transactions on Applied Perception* 2, 2 (2005), 106–117.
36. Roel Vertegaal. 2003. Attentive User Interfaces. *Commun. ACM* 46, 3 (2003), 31–33.
37. Jason Wither, Carmen E Au, Raymond Rischpater, and Radek Grzeszczuk. 2013. Moving beyond the map: Automated landmark based pedestrian guidance using street level panoramas. In *MobileHCI*. ACM, 203–212.
38. Qianli Xu, Liyuan Li, Joo Hwee Lim, Cheston Yin Chet Tan, Michal Mukawa, and Gang Wang. 2014. A Wearable Virtual Guide for Context-aware Cognitive Indoor Navigation. In *MobileHCI*. ACM, 111–120.