

Gaze-Guided Narratives: Adapting Audio Guide Content to Gaze in Virtual and Real Environments

Tiffany C.K. Kwok

Institute of Cartography and
Geoinformation, ETH Zürich
Zürich, Switzerland
ckwok@ethz.ch

Peter Kiefer

Institute of Cartography and
Geoinformation, ETH Zürich
Zürich, Switzerland
pekiefer@ethz.ch

Victor R. Schinazi

Chair of Cognitive Science,
ETH Zürich
Zürich, Switzerland
scvictor@ethz.ch

Benjamin Adams

Department of Geography,
University of Canterbury
Christchurch, New Zealand
benjamin.adams@canterbury.ac.nz

Martin Raubal

Institute of Cartography and
Geoinformation, ETH Zürich
Zürich, Switzerland
mraubal@ethz.ch

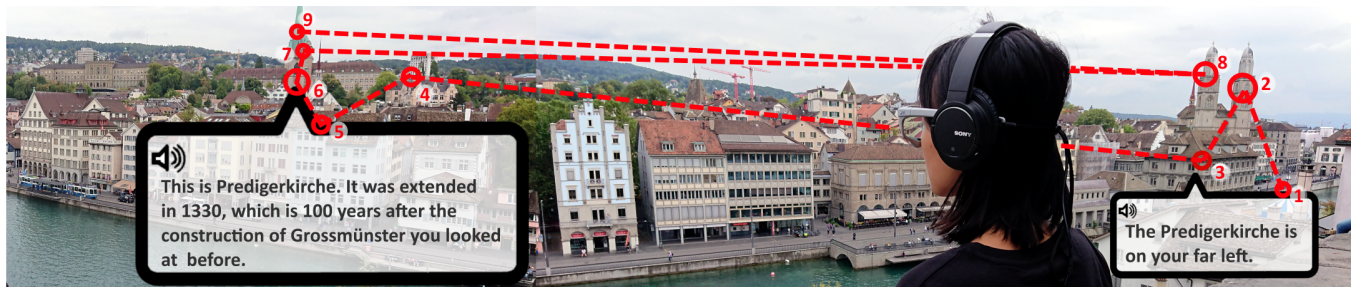


Figure 1: A tourist using the Gaze-Guided Narrative system from a vantage point (red: example fixation sequence with fixation locations (fixations 1 - 9)). The system provides gaze guidance when the user does not find the next object (fixation 3), and adapts the content to what has previously been looked at (fixation 6).

ABSTRACT

Exploring a city panorama from a vantage point is a popular tourist activity. Typical audio guides that support this activity are limited by their lack of responsiveness to user behavior and by the difficulty of matching audio descriptions to the panorama. These limitations can inhibit the acquisition of information and negatively affect user experience. This paper proposes *Gaze-Guided Narratives* as a novel interaction concept that helps tourists find specific features in the panorama (gaze guidance) while adapting the audio content to what has been previously looked at (content adaptation). Results from

a controlled study in a virtual environment ($n=60$) revealed that a system featuring both gaze guidance and content adaptation obtained better user experience, lower cognitive load, and led to better performance in a mapping task compared to a classic audio guide. A second study with tourists situated at a vantage point ($n=16$) further demonstrated the feasibility of this approach in the real world.

CCS CONCEPTS

• **Human-centered computing** → HCI theory, concepts and models; Usability testing; Field studies;

KEYWORDS

Gaze-Guided Narratives; Outdoor Eye Tracking; Tourist Guide

ACM Reference Format:

Tiffany C.K. Kwok, Peter Kiefer, Victor R. Schinazi, Benjamin Adams, and Martin Raubal. 2019. Gaze-Guided Narratives: Adapting Audio Guide Content to Gaze in Virtual and Real Environments. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019)*, May 4–9, 2019, Glasgow, Scotland Uk. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3290605.3300721>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHI 2019, May 4–9, 2019, Glasgow, Scotland Uk

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-5970-2/19/05...\$15.00

<https://doi.org/10.1145/3290605.3300721>

1 INTRODUCTION

Tourist destinations around the world have fascinating stories to tell and interesting facts to be revealed. Not surprisingly, tourism has always been among the early-adopting domains for novel HCI technologies [1, 5, 12, 18].

Audio guides are often used for telling stories about tourist destinations. Unlike visual displays, audio guides allow for an unobtrusive user experience by not distracting the tourist’s visual attention from the real world (see also [17]). In addition, audio can also be enriched to create immersive experiences including interactive multi-narrative soundscapes [42], 3D audio [6], or audio augmented reality [22].

A particular challenge of using audio, is that tourists must be able to map descriptions (e.g., “*the tower on your left*”) to the real world. This type of information mapping can lead to misunderstandings (due to ambiguity, cultural biases, or different conceptualizations of spatial descriptions) and to tourists getting lost while exploring a city’s panorama. Moreover, classic audio guides are only capable of playing at a specific pace and do not take into account the comprehension speed of their users. This lack of customization often generates confusion and can lead to cumbersome interactions that require stopping and rewinding the audio.

In this paper, we address these challenges with a gaze-based interaction concept called *Gaze-Guided Narratives*. *Gaze-Guided Narratives* adapt the audio content played to tourists exploring a panorama from a vantage point based on their real-time gaze. This technology is facilitated by recent progress in pervasive outdoor eye tracking [8, 19] that enables gaze-based interactions with objects in the wild [2] and developments in the smart glasses consumer market (e.g., the Pupil Labs eye tracking add-on for HoloLens [32]).

We believe that *Gaze-Guided Narratives* will help users find the reference objects (e.g., buildings) in a city panorama (*gaze guidance*), while adapting the audio content to objects that have been previously looked at (*content adaptation*). In a controlled lab study (n=60), we investigated the influence of gaze guidance and content adaptation on user experience, cognitive task load, content learning, and spatial learning for three different panoramas projected on a virtual CAVE (Cave Automatic Virtual Environment) environment. In a second study (n=16), we tested the *Gaze-Guided Narratives* interaction concept in the real world with tourists visiting a popular vantage point in Zürich, Switzerland.

Our contributions are:

- An implicit gaze-based interaction concept for audio guides, applied to touristic panorama views,
- An empirical evaluation of this interaction concept through a study in a controlled lab environment with 60 participants,

- A report of a real-world study demonstrating the feasibility of *Gaze-Guided Narratives* with 16 tourists in Zürich, Switzerland.

The paper is structured as follows. Section 2 positions our paper with regards to related work. In section 3, we introduce the *Gaze-Guided Narratives* concept and its implementation. Sections 4 and 5 describe and present the results of the lab and real-world studies. In section 6, we conclude the paper and provide an outlook for future work.

2 RELATED WORK

Tourist Guides

Researchers often use tourism as an opportunity to investigate context-aware pervasive or ubiquitous computing [1, 11]. Much of the early work in this field focused on location-based services for tourism including the creation of personalized city tours [40, 52]. Instead of guiding users through positioning technology, we investigate the interaction between users and a city panorama from a fixed location. Augmented reality (AR) applications have become a popular option for providing information about unknown locations in a city with the help of hand-held displays [35] or head-mounted displays [16]. However, researchers have highlighted the importance of unobtrusive interaction with nature [17]. Indeed, interacting with a display may separate users from the panorama and diminish their user experience (UX) by disconnecting them from the real world. Here, we focus on human environment interactions without a display that allow users to stay connected with the real world as they engage with a task or activity.

Pervasive Eye Tracking and Eye-Based Interaction

Eye-based interaction is commonly used in HCI because the user’s gaze can provide fast, natural, and intuitive ways to interact with different stimuli [36]. New developments in eye tracking technology have motivated researchers to work on pervasive eye tracking [8] in indoor [49] and outdoor [19] environments. Pervasive eye tracking has also become part of everyday life, found in gaze controlled computer games [53] and fatigue detection systems in cars [15]. Pervasive eye tracking can be conducted with web cams [41], smart-phone cameras [24], head-mounted eye trackers [26], with cameras installed in the environment [49], or below public displays [27]. Mobile eye trackers (such as the one used here) also enable hands-free interaction that allows users to focus on their surroundings. For example, Museum Guide 2.0 [55] enabled explicit gaze-based interaction with real 3D objects in an indoor space. Using Museum Guide 2.0 users could trigger events, which introduced the objects the users fixated on. Mobile eye tracking has also been used for studies on the way tourists explore (non-interactively) a city panorama [28]

as well as interact with a tourist map [44]. However, these systems did not allow users to interact directly with real 3D objects. To the best of our knowledge, there is no prior research on gaze-based interaction with a city panorama in the real world.

Different types of eye events, such as saccades, fixations [31], or smooth pursuits [50] can be used for eye-based interaction. Here, we use fixation data in order to determine the interaction. An interaction can be designed as either explicit or implicit [51]. With explicit interaction, users intentionally trigger the interaction. During implicit interaction, the system interprets the user’s regular interaction behavior on a higher level (e.g., in terms of activities [9]). We propose an implicit gaze-based interaction concept that adapts content based on previous visual attention and interprets user’s gaze in a wrong position as a failing visual search.

Gaze Guidance

The system we propose takes the role of a guide that helps the user find objects in the environment. Different approaches for gaze guidance have been suggested. Krejtz et al. [29] demonstrated that verbal audio descriptions can be used to guide a person’s gaze to a target. Gaze guidance has also been combined with vibrotactile feedback [25], visual feedback in AR [45], and non-verbal auditory feedback [34]. In addition, subconscious approaches to gaze guidance have been suggested that use image modulations in the user’s periphery [4, 37]. In this research, we use verbal audio descriptions, but make them dependent on the user’s current gaze position. Since the interpretations of verbal descriptions of spatial scenes often depend on the context [13], we conducted a pre-study (see section 3) to determine the parameters for our verbal gaze guidance.

Interactive Narratives and Spatial Narratives

The field of narratology has a history of investigating the structure of literary stories. Over the last decade, narratives and stories are increasingly being told through a variety of media. Indeed, the introduction of interactive narratives has led to the development of formal computational models of narrative [10]. Azaryahu and Foote [3] examined the way historical spaces can operate as a medium for telling *spatial narratives*. Looking at existing spatial narratives at historical touristic sites, these authors discuss the way the structure of narratives can vary from linear sequences in time and space to more complex non-linear configurations and configurations based on themes and sub-themes that operate over space and time. Although it is easier to construct linear (chronological) stories that maintain a dramatic arc, location-aware narrative systems can introduce non-linear structures depending on the amount of agency that users have to direct the storyline [14].

3 GAZE-GUIDED NARRATIVES

Concept

Traditional audio guides follow an information push paradigm. While many tourists appreciate the positive side of getting carefully selected content in a well-designed narrative sequence, traditional audio guides are somewhat inflexible. This can be problematic since the sequence of described objects is fixed and the audio speed does not often match the speed of comprehension and the activities taking place in the real world.

With the *Gaze-Guided Narratives* concept, we aim at diminishing some of the negative aspects of traditional audio guides. *Gaze-Guided Narratives* is based on two features:

1) *Gaze Guidance*: The system interactively helps the user find the objects of interest and only starts playing the content when the object has been found. Here, gaze guidance is provided together with directional cues (see section 2) in the form of verbal audio descriptions such as “*building A is far left of what you are looking at*” (refer to Figure 1, fixation 3). By providing gaze guidance along with directional cues, we prevent a break between the content parts and the guidance parts of the system while limiting the engagement of additional modalities (e.g., haptics) during interaction.

2) *Content Adaptation*: The system ensures flexibility regarding the sequence of objects described while maintaining the qualities of a good narrative by adapting the content to previously inspected objects (measured with eye tracking; refer to Figure 1, fixation 6). Content adaptations are made based on previously modeled relations between objects. In the context of tourism, we consider temporal, spatial, and thematic relations. For instance, if building B has been looked at at some point before building A, the system might tell that *A has been built before B* (temporal), *A is located left of B* (spatial), or *A and B have the same architectural style* (thematic). These kinds of adaptations should improve learning about the content and about the position of objects.

Implementation

Mapping Gaze to Objects in the Panorama. The system must be able to determine which object (e.g., building) in the panorama is being looked at. Our implementation is based on that suggested by [2]. A reference image is manually annotated with Areas of Interest (AOIs) around buildings that are relevant for the tourist guide. Geometric information of those AOIs is stored in the JSON format [2]. Objects are detected by matching the front-facing video of a head-mounted eye tracker to the annotated reference image using an ORB [46] feature detector. Fixations are calculated with the I-DT method [48] (thresholds: dispersion 1° , min duration 200ms). When a fixation in an AOI occurs an event is triggered and processed by the interaction logics. Apart from being robust

and fast for detection, this approach has been designed for stationary observers, such as tourists in our scenario.

Interaction Logics. Verbal directional instructions are provided by the system to help users identify buildings. Instructions are relative to the user’s current gaze (i.e., left, far left, right, far right, in front of, behind of) and are based on the thresholds determined by a pre-study (described in the next sub-section). A minimum time interval between each two consecutive instructions is needed in order to provide constant feedback to the user while ensuring that the user has sufficient time for interpreting and reacting to the instruction. We chose 4 seconds based on previous research on tactile navigation [43]. When the user successfully locates the objects in the panorama, the system starts to play the audio information. Based on the history of interaction, the system then decides which audio file (i.e., information with/without content adaptation) to play. While the audio content is playing, the user can explore the view without interruption.

Pre-study: Determining Parameters for Gaze Guidance. We performed a pre-study with 8 subjects in the CAVE to determine how participants interpret the verbal descriptions of directions used for gaze guidance when viewing a city panorama. Using the Zürich panorama as stimulus, participants were asked to fixate on a given starting location and then to fixate (head movements were allowed) on an ending location of their choice based on their interpretation of a directional instruction provided by the experimenter. Participants were tested on four directional instructions (left, far left, right, far right) with three starting locations (the center, the left, and the right of the city view). The vertical dimension (in front of, behind of) was not included in the pre-study. Table 1 presents the results of the saccade distances, which were then used to calculate the thresholds (mean plus one standard deviation) for the directional instructions used in the studies (see Table 2).

4 STUDY 1: CONTROLLED LAB STUDY

The goal of the first study was to evaluate the two main features (i.e., gaze guidance and content adaptation) of the *Gaze-Guided Narratives* system in relation to traditional audio guides. We designed four conditions in order to test the individual and interaction effects of these features (see Table 3). In condition A, the traditional audio guide serves as baseline. Conditions B and C include either the gaze guidance or the content adaptation features. Condition D simultaneously provides both gaze guidance and content adaptation.

This study focuses on the following research questions: **RQ1:** Does gaze guidance help participants identify and locate buildings? **RQ2:** Do Gaze-Guided Narratives improve the acquisition of information? **RQ3:** Do Gaze-Guided Narratives

enhance system usability and reduce cognitive load? **RQ4:** Do Gaze-Guided Narratives enhance user experience (UX)?

Methodology

Participants. Sixty participants (28 females) were recruited for the study. The age of the participants ranged from 18 to 55 years ($M = 23.3$, $SD = 6.4$). All participants were native German speakers and were not tourists at the time of the experiment. Participants were recruited via the DeSciL (Decision Science Lab) participant recruitment platform of ETH Zürich and were required to have normal or corrected to normal vision (with contact lenses) to participate.

Ethics statement. Written informed consent was obtained from all participants prior to starting the experiment. The participants were paid 30 CHF per hour and were told that they were allowed to abort the experiment at any time.

Materials. The experiment was conducted in a controlled CAVE environment with three large projection walls. A seat was placed at the center of the room at a distance of 1.8 meters from the front facing projection wall. The lights were turned off during the experiment. Participants wore SMI Eye Tracking Glasses (120Hz) and a Bluetooth SONY MDR-ZX770BN headphone during the experiment. The software modules provided by the eye tracking vendor were used for calibration and for recording the front-facing video image frames and raw gaze data.

Table 1: Pre-study for gaze guidance: Mean and standard deviations for saccade distances in visual angles for directional instructions.

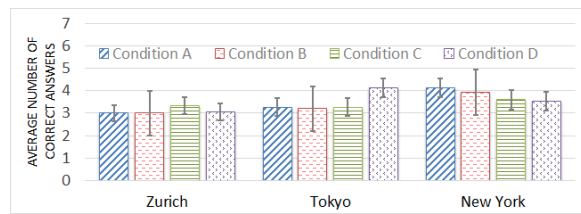
		Directional Instruction			
		Far Left	Left	Right	Far Right
Starting Location	Left	Not tested	-24.51° (-6.95°)	42.66° (25.22°)	62.43° (21.56°)
	Center	-55.44° (-14.64°)	-39.45° (-5.20°)	38.76° (6.35°)	58.47° (18.82°)
	Right	-61.53° (-13.85°)	-44.92° (-17.00°)	33.66° (7.57°)	Not tested

Table 2: Thresholds used for directional instructions in gaze guidance.

		Directional Instruction			
		Far Left	Left	Right	Far Right
Starting Location	Left	Not Used	< 0°	0° to 67.88°	> 67.88°
	Center	< -44.65°	-44.65° to 0°	0° to 45.11°	> 45.11°
	Right	< -61.92°	-61.92° to 0°	> 0°	Not Used

Table 3: The four conditions considered in the studies (Study 1: all conditions; Study 2: A and D).

		Content Adaptation	
		Without	With
Gaze Guidance	Without	Condition A Classic audio guide (baseline)	Condition B Adaptive narrative
	With	Condition C Gaze-guiding audio guide	Condition D Gaze-Guided Narrative

**Figure 2: Number of correctly answered content-related questions in study 1 (± standard error)**

We used panoramas from three different cities in order to vary the spatial structure and the types of buildings. The panoramas included a view from the “Tokyo Tower” in Tokyo, the “Top of the Rock” in New York, and the “Lindenhof” in Zürich. In order to account for differences in preferences and experience, we selected different themes for the audio content of the Tokyo (leisure and religion), New York (filming locations of famous movies) and Zürich (architecture) panoramas.

For each panorama, we selected four different buildings as target tourist attractions. The AOIs around these buildings were defined by polygons bounding the buildings without buffer. To ensure better control of the experiment, the sequence of AOIs was fixed and randomly generated before the creation of the audio content. The content for the different AOIs was created based on the official web-pages of each building (if available) and Wikipedia [57]. The content was chosen such that the audios for different AOIs were similar in length (approximately 1 minute). The scripts were originally written in English and translated into German. Audio files were generated by an online text-to-speech reader [38] and played through the headphone.

Participants completed the Santa Barbara Sense of Direction Scale (SBSODS) [21]. The SBSODS is a self-report measure of sense of direction ranging from zero (low) to seven (high). In order to measure cognitive load and system usability, participants completed the NASA Task Load Index (NASA-TLX) [20] and the System Usability Scale (SUS) [7], respectively. The scores for the NASA-TLX and SUS scores were calculated on a scale ranging from zero to one hundred.

Participants were also asked to complete a 6 factor User Experience Questionnaire (UEQ, [33]). This questionnaire measured the attractiveness (i.e., overall impression of the system), perspicuity (i.e., how easy is it to get familiar with the system), efficiency, dependability (i.e., whether users feel in control during the interaction), stimulation (i.e., if the user is excited and motivated to use the product) and novelty (i.e., the level of creativeness). Items in UEQ range from -3 to +3 with negative ratings representing a negative user experience.

At the end of each trial, participants completed a questionnaire with one multiple choice (e.g., “According to the audio, which of the following is the newest building?”) and six “true/false or unknown” (e.g., “Does Roppongi Hills Mori Tower have a museum in it?”) content-related questions specific to each panorama. Participants also completed a questionnaire (7-point Likert scale each) about their familiarity with the panorama and overall system helpfulness. Specifically, participants were asked whether the system was helpful in identifying the buildings, their familiarity with the cities presented in the experiment, their familiarity with the panorama they interacted with during the experiment and their familiarity with the content presented by the audio guide. All questionnaires were completed via paper and pencil.

Procedure. Upon arrival, participants were given an information sheet that described the experiment and were asked to sign the consent form. Participants then completed a short demographic questionnaire and the SBSODS before being asked to sit on the chair at the center of the CAVE. At this stage, the experimenter helped participants to adjust the eye tracker and headphones to their head. The eye tracker was 3-point calibrated at the beginning of each trial.

On each trial, the system started by familiarizing the participant with the sound effects for the audio transitions and the target buildings. For all conditions, the system first directed the participant’s gaze location to the target building with one initial instruction (see row “a” in Table 4). When transitioning from one building to another, exactly one description of the relative location to the previous building was provided (see row “d” in Table 4). Only the system with gaze guidance provided additional verbal feedback until the participant successfully located the building (refer to rows “b” and “e” in Table 4).

At the end of each trial, participants completed the UEQ, the content-related questions, and the questionnaire about familiarity and system helpfulness. In addition, participants were asked to draw a sketch map within an empty drawing canvas on a tablet. The canvas size had the same resolution as the image of the target environment. Subjects had to draw rectangles based on their memory for building locations. A

Table 4: Example for audio provided (with gaze guidance) and user actions for the Tokyo panorama.

System	User
a Now you can spend some time to look around. In the following part, you will hear information about the panorama of Tokyo... [...]... Now please look straight ahead. Zojoji Temple is on your right hand side.	Explores the panoramas freely and then tries to locate Zojoji Temple
b It is on the right of what you are looking at.... [more gaze-guidance until AOI found]	Searches and finds Zojoji Temple
c [sound effect for target found] Zojoji Temple .. [content]	Listens to the content
d We have now finished talking about Zojoji Temple. Now please look at Reiyukai Temple. You will find it to the far left of the Zojoji Temple.	Follows the instruction and search for the next building
e It is on the right of what you are looking at.... [more gaze-guidance until AOI found]	Searches and finds Reiyukai Temple
f	Continues to use the system
g This is the end of this part, please fill in the questionnaire provided.	Fills in post-session questionnaire

list with building names was provided during the sketch map task. After the last trial, participants filled in the NASA-TLX and the SUS questionnaires.

Study Design and Analysis. We adopted a between-subject design with participants randomly assigned to one of the four experimental conditions. Each participant completed three trials (one for each city panorama) in a pseudo-random order leading to 180 trials in total. The experiment required approximately 45 minutes to be completed. All statistical analyses were performed with SPSS.

Measurements. We were interested in how long it takes for participants to identify the building after the audio of the previous building has finished. The commonly used eye tracking measure time to first fixation [23] is not suitable in this case because the participant might have a random fixation on the building while still performing the search. Instead, we attempted to find the first fixation on the target AOI which falls into a phase of focal attention. We distinguished phases of focal and ambient attention using coefficient \mathcal{K} [30], which is calculated by subtracting the standardized (z-score) fixation duration from the standardized amplitude of the subsequent saccade. The *Time to Object Identification*

(TTOI) was defined as the time between the end of the last instruction (“d” in Table 4) and the first fixation on the target AOI in a phase of focal attention.

Sketch maps were analyzed using bidimensional regression (BDR) [54]. BDR quantifies the relationship between two sets of coordinates. The returned squared regression coefficient R^2 can be used as a similarity measure between the 2D configuration on the sketch map and that of the panorama image.

Results

Familiarity and Sense of Direction. Participants were more familiar with the Zürich panorama and its contents played through audio guide compared to New York and Tokyo. A one-way repeated-measures ANOVA revealed significant differences in the level of familiarity with the three panoramas ($F_{2,112} = 232.630, p < .001$). Post-hoc tests with Bonferroni correction further indicated that participants were more familiar with the Zürich panorama (6.9 ± 0.8) compared to the New York ($3.4 \pm 1.9, p < .001$) and the Tokyo ($1.6 \pm 1.4, p < .001$) panoramas. Participants were also more familiar with the New York panorama compared to Tokyo ($p < .001$). A one-way repeated-measures ANOVA also revealed significant differences in familiarity with the audio contents in different panoramas ($F_{2,112} = 55.839, p < .001$). Here again, post-hoc tests with Bonferroni correction showed that participants were more familiar with Zürich (3.3 ± 1.5) than New York ($2.6 \pm 1.5, p = .007$) and Tokyo ($1.4 \pm 0.9, p < .001$). Participants were also more familiar with the audio contents of New York compared to the contents of Tokyo ($p < .001$).

With regards to sense of direction, the mean SBSODS for all participants was 4.81 ($SD = 0.83$). Results from a one-way ANOVA revealed no significant differences in the self-reported sense of direction for participants in the four experimental conditions ($F_{3,56} = .492, p = .689$).

Content-Related Questions. After each trial, participants answered seven content-related questions. The total number of correctly answered questions in each trial was counted (see Figure 2). We conducted a 3 (city) x 4 (condition) mixed factorial ANOVA on the number of total correct answers across trials. Results revealed a significant main effect for city ($F_{2,112} = 18.26, p < .001$). Pairwise comparisons with Bonferroni correction further revealed that the average number of correct answers in Zürich (3.1 ± 1.4) and New York (3.6 ± 1.6) was lower compared to Tokyo (4.6 ± 1.6). There were no other main effects or interactions ($p > .59$).

Sketch Maps. A 3 (city) x 4 (condition) mixed factorial ANOVA on the results of the BDR (see example of a sketch map in Figure 3) revealed that there was a significant main effect of city ($F_{2,112} = 28.793, p < .001$) and condition ($F_{3,56} =$

3.852, $p < .014$). In addition, there was a significant interaction between city and condition ($F_{6,112} = 3.633, p = .015$). Additional pairwise contrasts with Bonferroni correction revealed that these differences were driven by the interaction between condition A (0.438 ± 0.416) and D (0.829 ± 0.310) for the Tokyo panorama ($p = .036$).

User Experience Ratings. The UEQ results are illustrated in Figure 4. In general, a trend can be observed in which the conditions with gaze guidance (C and D) were rated higher compared to the conditions without gaze guidance (A and B). We conducted a 3 (city) by 4 (condition) mixed ANOVA for each of the six factors in the UEQ questionnaire. Results revealed a significant main effect of city in terms of attractiveness ($F_{2,112} = 4.185, p = .018$), perspicuity ($F_{1,81,101.47} = 4.185, p < .001$), efficiency ($F_{2,112} = 7.331, p = .001$), dependability ($F_{2,112} = 17.240, p < .001$) and novelty ($F_{1,70,95.34} = 6.876, p = .002$). The between-subjects test revealed a significant main effect of condition for perspicuity ($F_{3,56} = 5.013, p = .004$) and dependability ($F_{3,56} = 4.070, p = .011$). Post-hoc tests with Bonferroni correction indicated that condition D had higher scores in perspicuity compared to conditions A ($p = .033$) and B ($p = .037$). Condition D also had higher scores in controllability when compared to condition B ($p = .024$). Results of the ANOVA also revealed a significant condition by city interaction for efficiency ($F_{6,112} = 2.208, p = .047$) and novelty ($F_{5,11,95.34} = 2.364, p = .035$). No significant difference was found in the post-hoc test analysis with Bonferroni correction.

System Usability and Cognitive Load. A one-way ANOVA comparing the four conditions in terms of system usability (SUS) revealed no significant differences between conditions (see Figure 5). With regards to cognitive load (NASA-TLX), results of a one-way ANOVA revealed a significant difference between conditions ($F_{3,56} = 3.752, p = .016$). Post-hoc tests with Bonferroni correction revealed that condition A had higher scores in cognitive load compared to condition D ($p = .035$). No significant differences were found between the other pairs of conditions.

Helpfulness in Identifying Building. Participants were asked in the post-trial questionnaire whether the system helped

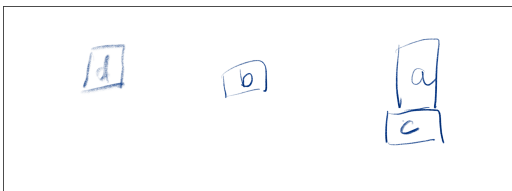


Figure 3: A sketch map drawn by a participant in study 1 (condition D, Zürich).

them to identify the buildings they were looking for. Results (see Figure 6) from a 3 (city) x 4 (condition) mixed ANOVA revealed a main effect of city ($F_{2,112} = 35.049, p < .001$) and condition ($F_{3,56} = 9.481, p < .001$). Results also revealed a significant condition by city interaction ($F_{6,112} = 5.310, p < .001$). Post-hoc tests with Bonferroni correction further indicated significant differences in the Tokyo panorama between conditions A and C ($p < .001$), conditions A and D ($p < .001$), conditions B and C ($p < .001$), and conditions B and D ($p = .023$). Significant pairwise differences were also found for the New York panorama between conditions A and C ($p = .014$), and conditions A and D ($p = .045$). No significant differences between conditions were found for the Zürich panorama.

Average Time to Object Identification. Figure 7 presents the average TTOI for the different conditions and cities. In general, conditions with gaze guidance tended to have longer TTOI than conditions without gaze guidance (i.e., condition C vs A and D vs B). Meanwhile, conditions with content adaptation also tended to have longer TTOI than that without content adaptation (i.e., condition B vs A and D vs C).

A 3 (city) x 4 (condition) mixed ANOVA revealed a main effect of city ($F_{2,112} = 11.799, p < .001$) and condition ($F_{3,56} = 3.572, p = .020$). Additional pairwise contrasts with Bonferroni correction revealed significant differences between New York ($13.94 \text{ s} \pm 1.232 \text{ s}$) and Zürich ($9.163 \text{ s} \pm 0.879 \text{ s}, p < .001$), and New York and Tokyo ($9.97 \text{ s} \pm 0.95 \text{ s}, p = .002$). Significant pairwise differences were also found between conditions D ($14.94 \text{ s} \pm 1.67 \text{ s}$) and A ($7.40 \text{ s} \pm 1.67 \text{ s}, p = .004$).

Discussion

Results from the experiment provide interesting insights on the interaction with *Gaze-Guided Narratives*. Critically, we found that:

RQ1: *Does gaze guidance help participants identify and locate buildings?* As expected, the conditions with gaze guidance (C and D) were rated as more helpful than conditions without (A and B). This result indicates that our approach to gaze guidance was accepted by users. Surprisingly, these effects were not apparent for the Zürich panorama.

Results of the eye tracking/TTOI analysis provide a more comprehensive understanding of the helpfulness of gaze guidance. Given that conditions with gaze guidance (C and D) require additional time for providing directional instructions, we did not expect to find differences in terms of TTOI when compared to conditions without gaze guidance (A and B). Findings were consistent with this expectation and revealed that it took approximately 7 seconds longer to identify the object in condition D compared to condition A.

Overall, gaze guidance appears to be more helpful in guiding users to identify buildings, albeit slower. In the context

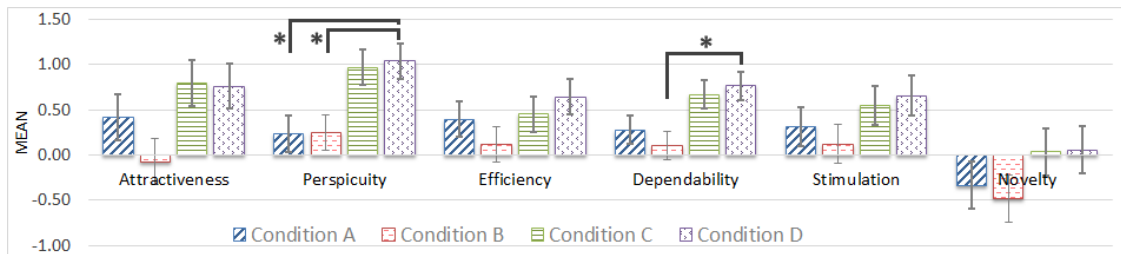


Figure 4: Comparison of UEQ factors between different conditions. Error bars indicate Standard Error, * significance.

of tourism, we believe that this delay is acceptable as tourists exploring a panorama are typically not under time pressure.

RQ2: Do Gaze-Guided Narratives improve the acquisition of information? Results from this study show that the number of correctly answered content-related questions did not significantly differ between different conditions. Interestingly, we found differences in correct answers between cities. Here, participants were better in answering these questions for Tokyo compared to New York and Zürich. We believe that these differences may be related to participants paying more attention to panoramas they were the least familiar with.

Results for the sketch map task revealed that participants were more accurate in drawing a sketch map of Zürich, followed by New York and then by Tokyo. This result was expected given their familiarity with these cities. However, sketch map accuracy was also significantly higher for condition D than condition A for the Tokyo panorama. This result is particularly interesting since it shows that *Gaze-Guided Narratives* is capable of assisting participants in the acquisition of spatial knowledge in unfamiliar cities.

Taken together, the results for the content-related questions and sketch maps, it seems that *Gaze-Guided Narratives* can be particularly helpful in unfamiliar panoramas – which is a common case in tourism.

RQ3: Do Gaze-Guided Narratives enhance system usability and reduce cognitive load? With gaze guidance, users were expected to have lower workload during search and more attention for listening to the audio guides. System usability was thus expected to improve by adding gaze guidance. Surprisingly, results revealed no significant differences in system usability between the four conditions. One reason for this may be that most of the participants complained that the text-to-speech voice used in the audio was unnatural. Since all conditions adopted the same text-to-speech, we suspect that negative effects caused by this limitation may counteract the potential positive effects of gaze guidance and content adaptation. In order to further investigate this issue, we used a native speaker to record the audios for the study in the real world (study 2, see section 5).

Although gaze guidance and content adaptation did not bring significant improvements in terms of system usability, they were capable of reducing the cognitive load when both features were present (condition D). This is an important finding since low cognitive load leaves free capacity that can be allocated for other tasks including the acquisition of information (see above, RQ2). Moreover, low cognitive workload means less fatigue and may allow tourists to stay attentive for longer periods of time.

RQ4: Do Gaze-Guided Narratives enhance user experience (UX)? Results revealed that condition D was easier to get familiar with than conditions A and B (perspicuity), and that it provided more controllability than condition B (dependability). The result on dependability was expected because gaze guidance does not overwhelm the user by pushing information in an uncontrolled speed.

A general trend can be observed in which condition D obtained higher scores compared to the baseline (condition A) in all six factors of the UEQ. However, no significant differences were found between the conditions in terms of attractiveness, efficiency, stimulation and novelty. Here again, this may be related to the quality of the text-to-speech voice in diminishing the positive effects brought on by the gaze guidance and content adaptation features of the system.

5 STUDY 2: REAL WORLD STUDY

We performed a study with tourists in order to investigate the feasibility of *Gaze-Guided Narratives* while exploring a city panorama in the real world (RQ5). In this experiment we focused solely on the classic audio guide (condition A) and the full *Gaze-Guided Narratives* system (condition D).

Methodology

Participants. Sixteen participants (8 females, $M = 30.1$, $SD = 6.5$, range 22-44) with normal or corrected to normal vision (with contact lenses) were recruited. All participants were native Chinese speaking tourists that were visiting Zürich for the first time and were not familiar with the testing site prior to the experiment. The average time spent by participants in Zürich prior to testing was 2.1 days ($SD = 1.4$; range 1 - 6).

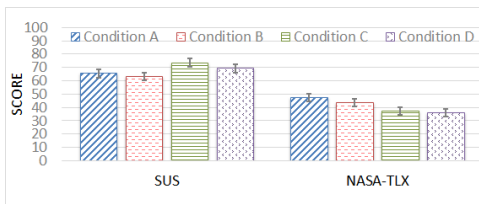


Figure 5: Average SUS and NASA-TLX scores in different conditions (\pm standard error).

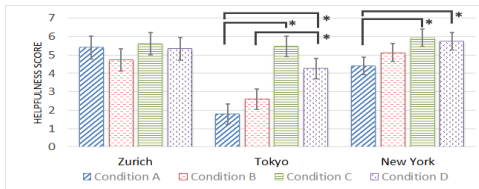


Figure 6: Average scores of helpfulness in identifying building in study 1 (\pm standard error), * significance.

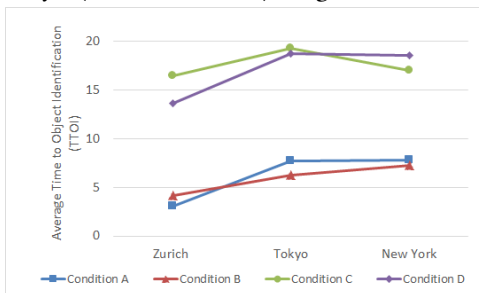


Figure 7: Average TTOI in different conditions and cities.

The decision to switch the language between study 1 and 2 was based on a 5-day observation at the experiment location that revealed that native Chinese speakers made up a large proportion of tourists arriving in small groups without a tourist guide.

Study Setup. The experiment was conducted at the Lindenhof in Zürich which provided participants with the same view as used in the Zürich stimulus (panorama) in study 1. Study 2 also used the same hardware and software for calibration and data collection as in study 1.

Materials. All materials from study 1 relevant for the Zürich panorama were translated into Chinese. Based on the feedback we received in study 1, the audio guides were recorded by a Chinese native speaker.

Procedure. Participants were recruited on site by the experimenter before they had a chance to inspect the vantage point. Potential participants were greeted and given a short explanation regarding the aim and procedure of the experiment. All participants completed the informed consent form prior to the start of the experiment. They were paid 20 CHF and were allowed to abort the experiment at any time.

The experiment procedure was similar to study 1 (see section 4) except that participants completed only one trial. In addition, the pre-study questionnaire included a question about the length of stay in Zürich, and the post-study questionnaire contained an additional question on whether the participants were interested in the content they listened to.

Study Design, Analysis and Measurements. We adopted a between subject design with participants randomly assigned to either condition A or D. Participants from both groups were exposed to the real-world panorama and required approximately 25 minutes to complete the experiment. All statistical analyses were performed with SPSS. Similar to experiment 1, participants completed the SBSODS, SUS and the NASA-TLX. Participants also answered content questions, familiarity questions and were asked to draw a sketch map. Sketch maps were analyzed using bidimensional regression and the average TTOI was calculated from the gaze data.

Results

Familiarity and Sense of Direction. Separate one-way ANOVAs were conducted to investigate differences between participants in conditions A and D in terms of their familiarity with the city of Zürich, their familiarity with the contents of the view, their level of interest with the contents in the view and their self-reported sense of direction (SBSDOS). Results revealed that participants in these two conditions did not differ in terms of familiarity with the city of Zürich ($F_{1,14} = 1.960, p = .183$), familiarity with the contents of the view ($F_{1,14} = 2.483, p = .137$), their interest in the contents of the view ($F_{1,14} = .360, p = .558$) and self-reported sense of direction ($F_{1,14} = .381, p = .547$).

User Experience Ratings. Table 5 presents results of the UEQ questionnaire. In general, a trend can be observed in which condition D obtains higher scores than condition A in all six user experience factors.

System Usability and Cognitive Load. The mean and standard deviation of SUS and NASA-TLX scores are presented in Table 6. The average SUS score for condition D ($M = 82.81, SD = 12.06$) is almost 15% higher than that of condition A ($M = 67.19, SD = 30.46$). The large standard deviation in the average SUS scores for condition A is related to an outlier who rated the SUS scores as 10.

Content Related Questions and Sketch Map. Similar to study 1, participants answered seven content-related questions and had to draw a sketch map without looking at the panorama. The number of correctly answered content-related questions and the R^2 values returned by BDR are presented in Table 6. Participants in condition D answered a higher number (4.13 questions) of correct questions and were more accurate (R^2 :

Table 5: Mean scores and standard deviations of UEQ factors in study 2 (-3/+3 = neg./pos. experience)

	Condition A		Condition D	
	M	(SD)	M	(SD)
Attractiveness	1.46	(0.77)	2.17	(0.87)
Perspicuity	1.72	(1.08)	2.19	(1.16)
Efficiency	1.03	(0.69)	1.88	(0.76)
Dependability	1.41	(1.03)	1.88	(1.16)
Stimulation	0.84	(0.74)	1.91	(0.95)
Novelty	0.94	(1.18)	2.19	(0.97)

Table 6: Mean and standard deviations of SUS (0 - 100), NASA-TLX scores (0 - 100), number of correct answers (0 - 7), BDR (R^2)(0 - 1) and TTOI (in sec) in study 2

	Condition A		Condition D	
	M	(SD)	M	(SD)
SUS	67.19	(30.46)	82.81	(12.06)
NASA-TLX	40.83	(15.77)	25.42	(12.92)
Number of Correct Answers	2.86	(1.36)	4.13	(2.10)
BDR (R^2)	0.88	(0.16)	0.96	(0.04)
Average TTOI	13.02	(17.87)	19.85	(14.37)

0.96) in drawing sketch maps compared to participants in condition A (2.86 questions and R^2 : 0.88).

Average Time to Object Identification. Results show that participants in condition A were faster (13.02 sec) compared to participants in condition D (19.85 sec) (see Table 6).

Open Comments from Participants. After the experiment, some participants expressed their positive attitude towards *Gaze Guided Narratives* by asking when it will be publicly available. Some of them suggested that a slower audio speed might have helped them capture important information.

Discussion

RQ5: Is it feasible to use *Gaze-Guided Narratives* in the real world? Comparing the results on system usability between the two studies, we find that the system with *Gaze-Guided Narratives* (condition D) obtained an average SUS score of 82.81 (SD = 12.06) in the real world and 69.33 (SD = 12.31) in the CAVE (see section 4). An average SUS score above 80 has been suggested to be considered as “pretty good”, while a score of 66 could be considered as average [56]. One possible explanation for this difference could be the higher degree of immersion in the real world. It is also possible that replacing text-to-speech with a voice recording had an influence on system usability. Overall, the SUS score results indicate that participants were capable of using the system in the real world without loss in system usability.

Inferential statistics were not conducted in study 2 because of the limited number of participants (16 participants). However, results for the NASA-TLX, UEQ and TTOI are consistent with those obtained in study 1. Together, these findings suggest that applications of *Gaze-Guided Narratives* are not limited to virtual environments. Interestingly, study 2 obtained higher scores for all six factors in UEQ, for both condition A and D. This general trend may also be caused by the introduction of a more natural voice in the audio.

Missing Gaze Data. The challenge of achieving a high tracking quality has always been considered as a core research question in pervasive eye tracking [8] and eye tracking “in the wild” [19]. Although the tracking quality was not our main research focus and often determined by the commercial eye tracking hardware, the feasibility of using our system in the real world (**RQ5**) may be nevertheless influenced by it. Sunlight can disturb the reflective properties of the infrared light used by the eye tracker. In the CAVE environment with controlled lighting, the average percentage of missing gaze data for 180 trials was 2.14%, which is 9.66% less than that in the outdoor environment for 16 trials of data (11.8%). Although the percentage of missing gaze data in the real world was larger than that in a controlled environment, this did not seem to affect the SUS scores, and no participant complained about interruptions or unresponsive system behavior.

6 CONCLUSION AND OUTLOOK

We have proposed *Gaze-Guided Narratives* as an implicit gaze-based interaction concept for audio guides, which is particularly suited for touristic panorama views. The concept was evaluated through an empirical controlled lab study with 60 participants. Results revealed that the *Gaze-Guided Narratives* system obtained better UX, lower cognitive load, and better performance in a mapping task compared to a classic audio guide. We demonstrated the feasibility of our approach “in the wild” with a real-world study with 16 tourists.

Although the results of our study are promising, they represent only the start of gaze-based interaction for tourist assistance. Next steps include gaze-based tourist recommendations, interest detection, and a support of touristic activities beyond panorama exploration including wayfinding or shopping. These directions will be accelerated by further developments and pervasive eye tracking technology. Finally, *Gaze-Guided Narratives* could be applied beyond tourism scenarios, such as for mobile learning [47] of place-related content or for the creation of dynamic story lines in location-based games [39].

ACKNOWLEDGMENTS

This work is supported by an ETH Zürich Research Grant [ETH-38 14-2].

REFERENCES

- [1] Gregory D Abowd, Christopher G Atkeson, Jason Hong, Sue Long, Rob Kooper, and Mike Pinkerton. 1997. Cyberguide: A mobile context-aware tour guide. *Wireless networks* 3, 5 (1997), 421–433.
- [2] Vasileios-Athanasios Anagnostopoulos, Michal Havlena, Peter Kiefer, Ioannis Giannopoulos, Konrad Schindler, and Martin Raubal. 2017. Gaze-Informed Location Based Services. *International Journal of Geographical Information Science* 31, 9 (2017), 1770–1797. <https://doi.org/10.1080/13658816.2017.1334896>
- [3] Maoz Azaryahu and Kenneth E. Foote. 2008. Historical space as narrative medium: on the configuration of spatial narratives of time at historical sites. *GeoJournal* 73, 3 (Nov 2008), 179–194. <https://doi.org/10.1007/s10708-008-9202-4>
- [4] Reynold Bailey, Ann McNamara, Nisha Sudarsanam, and Cindy Grimm. 2009. Subtle gaze direction. *ACM Transactions on Graphics (TOG)* 28, 4 (2009), 100.
- [5] Rafael Ballagas, André Kuntze, and Steffen P Walz. 2008. Gaming tourism: Lessons from evaluating REXplorer, a pervasive game for tourists. In *International Conference on Pervasive Computing*. Springer, 244–261.
- [6] Stephen Brewster, Joanna Lumsden, Marek Bell, Malcolm Hall, and Stuart Tasker. 2003. Multimodal 'eyes-free' interaction techniques for wearable devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 473–480.
- [7] John Brooke. 1996. SUS - A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7. <https://doi.org/10.1002/hbm.20701>
- [8] Andreas Bulling, Andrew T Duchowski, and Päivi Majaranta. 2011. PETMEI 2011: the 1st international workshop on pervasive eye tracking and mobile eye-based interaction. In *Proceedings of the 13th international conference on Ubiquitous computing*. ACM, 627–628.
- [9] Andreas Bulling, Jamie A. Ward, Hans Gellersen, and Gerhard Troster. 2011. Eye Movement Analysis for Activity Recognition Using Electrooculography. *IEEE Trans. Pattern Anal. Mach. Intell.* 33, 4 (April 2011), 741–753. <https://doi.org/10.1109/TPAMI.2010.86>
- [10] Marc Cavazza and David Pizzi. 2006. Narratology for interactive storytelling: A critical introduction. In *International Conference on Technologies for Interactive Digital Storytelling and Entertainment*. Springer, 72–83.
- [11] Keith Cheverst, Nigel Davies, Keith Mitchell, and Adrian Friday. 2000. Experiences of developing and deploying a context-aware tourist guide: the GUIDE project. In *Proceedings of the 6th annual international conference on Mobile computing and networking*. ACM, 20–31.
- [12] Keith William John Cheverst, Ian Norman Gregory, and Helen Turner. 2016. Encouraging Visitor Engagement and Reflection with the Landscape of the English Lake District: Exploring the potential of Locative Media. In *International Workshop on Unobtrusive User Experiences with Technology in Nature*. 1–5.
- [13] Kenny R Coventry, Thora Tenbrink, and John Bateman. 2009. *Spatial language and dialogue*. Vol. 3. OUP Oxford.
- [14] Steven Dow, Jaemin Lee, Christopher Oezbek, Blair MacIntyre, Jay David Bolter, and Maribeth Gandy. 2005. Exploring spatial narratives and mixed reality experiences in Oakland Cemetery. In *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology*. ACM, 51–60.
- [15] Lex Fridman, Heishiro Toyoda, Sean Seaman, Bobbie Seppelt, Linda Angell, Joonbum Lee, Bruce Mehler, and Bryan Reimer. 2017. What Can Be Predicted from Six Seconds of Driver Glances?. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 2805–2813. <https://doi.org/10.1145/3025453.3025929>
- [16] Brian F. Goldiez, Ali M. Ahmad, and Peter A. Hancock. 2007. Effects of Augmented Reality Display Settings on Human Wayfinding Performance. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 37, 5 (Sept 2007), 839–845. <https://doi.org/10.1109/TSMCC.2007.900665>
- [17] Jonna Häkkinilä, Keith Cheverst, Johannes Schöning, Nicola J. Bidwell, Simon Robinson, and Ashley Colley. 2016. NatureCHI: Unobtrusive User Experiences with Technology in Nature. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16)*. ACM, New York, NY, USA, 3574–3580. <https://doi.org/10.1145/2851581.2856495>
- [18] Dai-In Han, Timothy Jung, and Alex Gibson. 2013. Dublin AR: implementing augmented reality in tourism. In *Information and communication technologies in tourism 2014*. Springer, 511–523.
- [19] Dan Witzner Hansen and Arthur E.C. Pece. 2005. Eye tracking in the wild. *Computer Vision and Image Understanding* 98, 1 (2005), 155 – 181. <https://doi.org/10.1016/j.cviu.2004.07.013> Special Issue on Eye Detection and Tracking.
- [20] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, Peter A. Hancock and Najmedin Meshkati (Eds.). Advances in Psychology, Vol. 52. North-Holland, 139 – 183. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)
- [21] Mary Hegarty. 2002. Development of a self-report measure of environmental spatial ability. *Intelligence* 30, 5 (2002), 425–447. [https://doi.org/10.1016/S0160-2896\(02\)00116-2](https://doi.org/10.1016/S0160-2896(02)00116-2)
- [22] Florian Heller, Aaron Krämer, and Jan Borchers. 2014. Simplifying Orientation Measurement for Mobile Audio Augmented Reality Applications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 615–624. <https://doi.org/10.1145/2556288.2557021>
- [23] Kenneth Holmqvist, Marcus Nyström, Richard Andersson, Richard DeWhurst, Halszka Jarodzka, and Joost Van de Weijer. 2011. *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.
- [24] Michael Xuelin Huang, Jijia Li, Grace Ngai, and Hong Va Leong. 2017. ScreenGlint: Practical, In-situ Gaze Estimation on Smartphones. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 2546–2557. <https://doi.org/10.1145/3025453.3025794>
- [25] Jari Kangas, Deepak Akkil, Jussi Rantala, Poika Isokoski, Päivi Majaranta, and Roope Raisamo. 2014. Gaze Gestures and Haptic Feedback in Mobile Devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 435–438. <https://doi.org/10.1145/2556288.2557040>
- [26] Moritz Kassner, William Patera, and Andreas Bulling. 2014. Pupil: An Open Source Platform for Pervasive Eye Tracking and Mobile Gaze-based Interaction. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication (UbiComp '14 Adjunct)*. ACM, New York, NY, USA, 1151–1160. <https://doi.org/10.1145/2638728.2641695>
- [27] Mohamed Khamis, Axel Hoesl, Alexander Klimczak, Martin Reiss, Florian Alt, and Andreas Bulling. 2017. EyeScout: Active Eye Tracking for Position and Movement Independent Gaze Interaction with Large Public Displays. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology (UIST '17)*. ACM, New York, NY, USA, 155–166. <https://doi.org/10.1145/3126594.3126630>
- [28] Peter Kiefer, Ioannis Giannopoulos, Dominik Kremer, Christoph Schlieder, and Martin Raubal. 2014. Starting to get bored: An outdoor eye tracking study of tourists exploring a city panorama. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '14)*. ACM, New York, NY, USA, 315–318. <https://doi.org/10.1145/2578153.2578216>

- [29] Izabela Krejtz, Agnieszka Szarkowska, Krzysztof Krejtz, Agnieszka Walczak, and Andrew Duchowski. 2012. Audio Description As an Aural Guide of Children’s Visual Attention: Evidence from an Eye-tracking Study. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA ’12)*. ACM, New York, NY, USA, 99–106. <https://doi.org/10.1145/2168556.2168572>
- [30] Krzysztof Krejtz, Andrew Duchowski, Izabela Krejtz, Agnieszka Szarkowska, and Agata Kopacz. 2016. Discerning Ambient/Focal Attention with Coefficient K. *ACM Trans. Appl. Percept.* 13, 3, Article 11 (May 2016), 20 pages. <https://doi.org/10.1145/2896452>
- [31] Kuno Kurzhals, Emine Cetinkaya, Yongtao Hu, Wenping Wang, and Daniel Weiskopf. 2017. Close to the Action: Eye-Tracking Evaluation of Speaker-Following Subtitles. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI ’17)*. ACM, New York, NY, USA, 6559–6568. <https://doi.org/10.1145/3025453.3025772>
- [32] Pupil Labs. 2017. Hololens and BT300 eye tracking add-ons. <https://pupil-labs.com/blog/2017-03/hololens-and-bt300-eye-tracking-add-ons/>. Accessed: 2018-12-25.
- [33] Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and Evaluation of a User Experience Questionnaire. In *HCI and Usability for Education and Work*, Andreas Holzinger (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 63–76.
- [34] Viktor Losing, Lukas Rottkamp, Michael Zeunert, and Thies Pfeiffer. 2014. Guiding visual search tasks using gaze-contingent auditory feedback. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. ACM, 1093–1102.
- [35] Patrick Luley, Roland Perko, Johannes Weinzerl, Lucas Paletta, and Alexander Almer. 2012. Mobile Augmented Reality for Tourists - MARFT. *Advances in Location-Based Services Lecture Notes in Geoinformation and Cartography* (2012), 21–36. https://doi.org/10.1007/978-3-642-24198-7_2
- [36] Päivi Majaranta and Andreas Bulling. 2014. *Eye Tracking and Eye-Based Human-Computer Interaction*. Springer London, London, 39–65. https://doi.org/10.1007/978-1-4471-6392-3_3
- [37] Ann McNamara, Thomas Booth, Srinivas Sridharan, Stephen Caffey, Cindy Grimm, and Reynold Bailey. 2012. Directing gaze in narrative art. In *Proceedings of the ACM Symposium on Applied Perception*. ACM, 63–70.
- [38] NaturalSoft Ltd. 2018. NaturalReader Commercial. <https://www.naturalreaders.com/commercial.html> [Online].
- [39] Marcel Neuenhaus and Maha Aly. 2017. Empathy Up. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA ’17)*. ACM, New York, NY, USA, 86–92. <https://doi.org/10.1145/3027063.3049276>
- [40] Valeria Orso, Alessandra Varotto, Stefano Rodaro, Anna Spagnolli, Giulio Jacucci, Salvatore Andolina, Jukka Leino, and Luciano Gamberini. 2017. A Two-step, User-centered Approach to Personalized Tourist Recommendations. In *Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter (CHIItaly ’17)*. ACM, New York, NY, USA, Article 7, 5 pages. <https://doi.org/10.1145/3125571.3125594>
- [41] Seonwook Park, Xucong Zhang, Andreas Bulling, and Otmar Hilliges. 2018. Learning to Find Eye Region Landmarks for Remote Gaze Estimation in Unconstrained Settings. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications (ETRA ’18)*. ACM, New York, NY, USA, Article 21, 10 pages. <https://doi.org/10.1145/3204493.3204545>
- [42] Daniela Petrelli, Nick Dulake, Mark T. Marshall, Anna Pisetti, and Elena Not. 2016. Voices from the War: Design As a Means of Understanding the Experience of Visiting Heritage. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI ’16)*. ACM, New York, NY, USA, 1033–1044. <https://doi.org/10.1145/2858036.2858287>
- [43] Martin Pielot, Benjamin Poppinga, Wilko Heuten, and Susanne Boll. 2012. PocketNavigator: Studying Tactile Navigation Systems In-situ. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’12)*. ACM, New York, NY, USA, 3131–3140. <https://doi.org/10.1145/2207676.2208728>
- [44] Pernilla Qvarfordt, David Beymer, and Shumin Zhai. 2005. RealTourist – A Study of Augmenting Human-Human and Human-Computer Dialogue with Eye-Gaze Overlay. In *Human-Computer Interaction - INTERACT 2005*, Maria Francesca Costabile and Fabio Paternò (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 767–780.
- [45] Patrick Renner and Thies 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. <https://doi.org/10.1109/3DUI.2017.7893338>
- [46] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. 2011. ORB: An efficient alternative to SIFT or SURF. In *2011 International Conference on Computer Vision*. 2564–2571. <https://doi.org/10.1109/ICCV.2011.6126544>
- [47] Christian Sailer, Peter Kiefer, and Raubal Martin. 2015. An integrated learning management system for location-based mobile learning. *Proceedings of the 11th International Conference on Mobile Learning 2015*, 118–122.
- [48] Dario D. Salvucci and Joseph H. Goldberg. 2000. Identifying Fixations and Saccades in Eye-tracking Protocols. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications (ETRA ’00)*. ACM, New York, NY, USA, 71–78. <https://doi.org/10.1145/355017.355028>
- [49] Thiago Santini, Hanna Brinkmann, Luise Reitstätter, Helmut Leder, Raphael Rosenberg, Wolfgang Rosenstiel, and Enkelejda Kasneci. 2018. The art of pervasive eye tracking: unconstrained eye tracking in the Austrian Gallery Belvedere. In *Proceedings of the 7th Workshop on Pervasive Eye Tracking and Mobile Eye-Based Interaction*. ACM, 5.
- [50] Simon Schenk, Marc Dreiser, Gerhard Rigoll, and Michael Dorr. 2017. GazeEverywhere: Enabling Gaze-only User Interaction on an Unmodified Desktop PC in Everyday Scenarios. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI ’17)*. ACM, New York, NY, USA, 3034–3044. <https://doi.org/10.1145/3025453.3025455>
- [51] Albrecht Schmidt. 2000. Implicit human computer interaction through context. *Personal Technologies* 4, 2 (01 Jun 2000), 191–199. <https://doi.org/10.1007/BF01324126>
- [52] Johannes Schöning, Brent Hecht, and Nicole Starosielski. 2008. Evaluating automatically generated location-based stories for tourists. In *CHI’08 extended abstracts on Human factors in computing systems*. ACM, 2937–2942.
- [53] Tobii. 2018. Enhanced PC Games with Eye Tracking. <https://tobiigaming.com/games/>. Accessed: 2018-12-25.
- [54] Waldo R. Tobler. 1994. Bidimensional Regression. *Geographical Analysis* 26, 3 (1994), 187–212. <https://doi.org/10.1111/j.1538-4632.1994.tb00320.x>
- [55] Takumi Toyama, Thomas Kieninger, Faisal Shafait, and Andreas Dengel. 2011. Museum Guide 2.0 – An Eye-Tracking based Personal Assistant for Museums and Exhibits. 1 (05 2011), 103–110.
- [56] Tom Tullis and Bill Albert. 2013. Chapter 6 - Self-Reported Metrics. In *Measuring the User Experience (Second Edition)* (second edition ed.), Tom Tullis and Bill Albert (Eds.). Morgan Kaufmann, Boston, 121 – 161. <https://doi.org/10.1016/B978-0-12-415781-1.00006-6>
- [57] Wikimedia Foundation, Inc. 2018. Wikipedia: The free encyclopedia. <https://www.wikipedia.org> [Online].