Interweaving Visual and Audio-Haptic Augmented Reality for Urban Exploration

Yi-Ta Hsieh[1], **Valeria Orso**[2], **Salvatore Andolina**[3], **Manuela Canaveras**[2], **Diogo Cabral**[4], **Anna Spagnolli**[2], **Luciano Gamberini**[2], **Giulio Jacucci**[1,3]
[1]Helsinki Institute for Information Technology HIIT, Dept. Computer Science, Univ. Helsinki, Finland
first.last@helsinki.fi
[2]Human Inspired Technology Research Centre HIT, Dept. General Psychology, Univ. Padova, Italy
first.last@unipd.it
[3]Helsinki Institute for Information Technology HIIT, Aalto University, Finland
first.last@aalto.fi
[4]Madeira-ITI, University of Madeira, Portugal
first.last@m-iti.org

ABSTRACT
While ordinary touchscreen-based interfaces on urban explorer applications draw much of a user's attention onto the screen, visual and audio-haptic augmented reality interfaces have emerged as the two main streams for enabling direct focus on the surroundings. However, neither interface alone satisfies users in the highly dynamic urban environment. This research investigates how the two complementary augmentation can coexist on one system and how people adapt to the situation by selecting the more suitable interface. A prototype was deployed in a field experiment in which participants explored points of interest in an urban environment with both interfaces. The engagement with the surroundings was compared with a touchscreen-based application. Most participants spontaneously switched between the two interfaces, which manifests the value of the availability of both interfaces on one system. The results point at the situated advantages of either interface and reveal the users' preferences when both interfaces are available.

ACM Classification Keywords
H.5.2. Information Interfaces and Presentation (e.g. HCI): User Interfaces

Author Keywords
Augmented Reality; Multimodal Interaction; Urban Exploration; Audio-Haptic Interface.

DIS 2018, June 9–13, 2018, Hong Kong.

ACM ISBN 978-1-4503-5198-0/18/06 ...$15.00.
http://dx.doi.org/10.1145/3196709.3196733

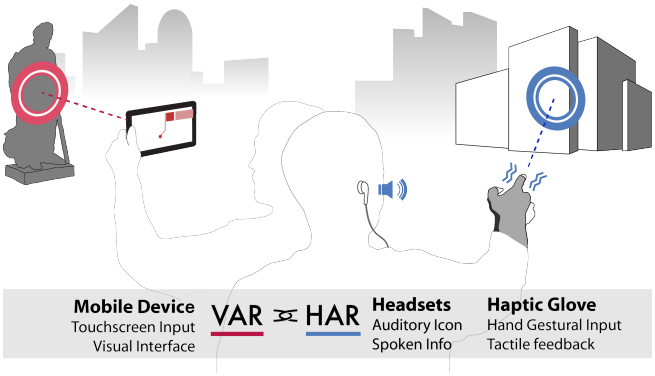

Figure 1. The availability of both VAR and HAR interfaces on one urban heritage explorer enables users to freely choose which interface to use to easily adapt to the highly dynamic environment while maintaining focus on the environment.

INTRODUCTION
This paper investigates how visual augmented reality (VAR) and audio-haptic augmented reality (HAR) interfaces in urban exploration can support users in focusing on the real content—the situated surroundings—rather than on the interface itself, as well as how these two interfaces can coexist on one system to support adaptation to the dynamic environment. It is very common that mobile users tend to focus on interacting with the touchscreen-equipped device, and such head-down interaction results in not only ignoring the surroundings but also increasing unsafe behaviors [11, 27, 15]. This could severely hamper the experience in which the surroundings are the actual content to interact with and the environment is highly dynamic.

Two main streams of augmented reality (AR) solutions have emerged from the interaction design point of view: sensor-based VAR and non-visual HAR approaches. Both proposals adopt the general idea of aligning users' pointing direction, the device, and real-world objects to enable direct focus on

the environment. The principle of VAR approach is to utilize on-device sensors to understand users' viewing orientation so that geo-referenced digital content can be overlaid onto the camera preview of the real scene. For example, Wikitude[1] enables users to point the device in a particular direction to look for information associated with objects in the situated surroundings. Similarly, HAR solutions have mainly deployed handheld devices as a "magic wand" to point directly at specific landmarks for fetching associated information [8]. This approach overcomes the requirement of concentrating users' visual attention on the interface, allowing their visual resources to be fully allocated to the surroundings.

Although both VAR and HAR interfaces can enable direct focus on the surroundings, each has specific strengths and weaknesses in different contexts of urban exploration. For instance, when a user's priority is to search for a specific piece of information (e.g., searching the stories about a statue by its name), a VAR interface allows him/her to quickly skim through the contents. On the other hand, when users need to keep their eyes on the environment (e.g., in rush hour by a road), HAR interfaces might serve users better. Urban exploration tools should meet users' dynamic and evolving needs even by supporting the switch between different modalities [37]. This motivates us to investigate how the two complementary interfaces can coexist on one system.

Our aim is to gain insights into designing an urban exploration system equipped with both VAR and HAR interfaces. Since grounding knowledge to design such an integrated system is still unknown, we identified two fundamental aspects to be studied. The first aspect entails identifying the circumstances in which each interface would benefit users the most, in order to further improve the design in each area. Secondly, we aim at understanding how a smooth transition can be designed when switching from one interface to the other. If the effort in the interface switching is too high, users would rather continue using the less suitable interface than switching to the other.

Our research method is to build an urban explorer prototype interwoven with both VAR and HAR interfaces as a research tool and deploy it in a field experiment in which participants explore and locate urban points of interest (POIs). While the pointing techniques in sensor-based VAR interfaces have been well established with various applications available on the market, the interaction techniques in a HAR interface vary from one form factor of the sensing device to another. Since handheld devices occupy the hand, many wearable sensing devices have been developed for eliminating constraints on the freedom of the hand, such as a belt [34], a wristband [12], or a ring [16]. Considering the benefits of leveraging our already-familiar hand gestures to directly interact with landmarks, we adopted a spatial sensing haptic glove (SSH Glove) for the HAR interface. The glove features hand orientation and gestural sensing, as well as providing tactile feedback. As a result, the prototype system incorporates a mobile device, a glove, and a headset (See Fig. 1). The research prototype may seem bulky, but it serves well at this exploratory stage.

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

We report an experiment in which 18 participants used our prototype to explore the central area of a historic city in the context of public, real-world use. In the three-phase experiment, the participants were allowed to first use VAR and HAR interfaces separately, and then to freely choose and switch between VAR and HAR during the last phase of exploration. We observed spontaneous switching of interfaces from most participants during the exploration, which manifests the value of the availability of both interfaces on one system. Moreover, to reveal the benefits of an urban explorer with both VAR and HAR interfaces, a control group of 18 participants used the Google Maps application to explore the same area. Results indicate that it is possible to predict the way people would use those modes in different situations. Knowledge gained from this research fosters an ultimate visual-audio-haptic AR interface for urban exploration, and could direct further studies that are carried out in the field with realistic tasks but a controlled design.

BACKGROUND

Mobile scenarios have been the favorite research setting to explore new interfaces that extend and augment human capabilities [3] with visual, haptic, and aural augmentations. Such multimodal interaction methods aim at improving human-computer communications by utilizing all available modalities of human input and output in a natural manner [36].

Exploring Surroundings with Visual AR

Based on the "smart lens" technique, VAR interfaces allow users to "see through" a device via the camera preview, hence allowing direct focus on the environment [8]. VAR applications also benefit users by providing information relevant to the current location, enabling access to timely and updated multimedia content and supporting interactive annotations [41, 23]. These multimedia augmentations can provide an immersed experience in the immediate environment. In addition, VAR exploration can increase the chances of finding relevant information, as noted by Schmalstieg and Höllerer [32].

VAR applications have been developed for increasing users' awareness of their surroundings [4, 6] but with certain practical constraints when in use. For example, its outdoor usage presents additional interaction challenges. Schmalstieg and Höllerer [32] state that most of the displays are not bright enough to achieve sufficient contrast in outdoor situations and mobile devices have a bigger risk of creating fatigue, particularly when one has to hold them at eye level. In addition to these issues, there is also the difficulty of touching on a display while holding it at the same time [28] and the difficulty of reading text on small size displays [7, 18]. Such interaction issues could potentially be avoided when audio and haptic modalities are incorporated in the interaction.

Exploring Surroundings with Audio-Haptic AR

Audio and haptic modalities can be incorporated into the interface design so that users' visual resource can be focused even more on the surroundings. McGookin et al. [26] presented Audio Bubbles that associate audio with real-world landmarks to support serendipitous discovery in tourist activities. By overlaying virtual audio in the soundscape of the environment,

their approach can increase the awareness of nearby interesting things. Similarly, auditory icons and music icons in the field can support serendipitous discovery, emotional engagement, and implicit navigation [1, 21, 39]. Audio can be further complemented with haptic modality. Combining the two for displaying touristic information results in a higher recognition rate than utilizing each modality alone [29].

Moreover, using the handheld device as a "magic wand" to point directly at real-world objects imposes an even more explicit connection with the surroundings [8]. For example, Magnusson et al. [25] presented an audio-tactile tourist guide on a mobile device through which users can scan for the direction of POIs and receive audio-recorded information. This point-and-select type of interaction is perceived as easy to use and natural [25, 31]. However, it comes at a cost of carrying a device that limits the freedom of users' hands. Moreover, attention is soon drawn back to the device after the selection, rather than remaining on the surroundings [8].

The constraint on the freedom of the hand can be easily confronted by deploying hand-worn form factors on the device while still enjoying the inherited benefits from the "magic wand" technique. Moreover, we wanted to leverage familiar gestures (e.g., finger pointing or grabbing gesture) to construct a stronger augmented experience and connection with the surroundings. As compared to other hand-worn form factors, such as wristbands [12] or rings [16], the glove form factor affords sufficient space for distributing sensors on various locations of the hand for enabling more truthful hand gestural recognition; hence, a more complete gestural language set is possible. For example, Myopoint, a sensor-equipped wristband, uses a rather counterintuitive gesture—opening one's palm—for selection, possibly due to limited capability in detecting finer finger gestures [12]. There are other wearable form factor options, such as a belt for wayfinding [34, 30]. However, they do not support magic wand-type pointing interaction. Hence, a glove was more suitable than belts or vests. Another benefit of using the glove form factor is to provide tactile feedback on an action-performing body. In sum, using the glove can leverage our existing skills in using the hand to interact with everyday objects, resulting in a more intimate connection with the surroundings through the interaction.

However, delivering information over audio-haptic interfaces could be less efficient than through visual interfaces. At least, the duration of the spoken content is the minimal time needed from users to receive the complete information. Although fast forward is possible on audio interfaces, it is still not as efficient as skimming text on a visual interface. This fact further ascertains how VAR and HAR interfaces could complement each other.

Other Approaches

Vaittinen and McGookin [37] point out that most of current tourism applications are based on maps and lists (e.g., Google Maps[2] and Tripadvisor[3]), and that do not work for all needs, particularly for touristic/urban exploration. The applications would benefit from interfaces that could register the information spatially to the objects, removing the need to divert the focus to the device. Baldauf et al. [2] have compared different (but traditional) visualization techniques of geo-referenced information for supporting mobile urban exploration. They have compared 2D and 3D maps, list and category views, and tag clouds. In their study, the users preferred 2D maps and category view to tag view and 3D maps. However, they missed an opportunity to study VAR and audio-haptic interfaces.

Equipped with see-through displays, smart glasses allow users to view both digital content and a real scene simultaneously, which potentially can be an ideal solution for enabling direct focus on the environment and increasing situational awareness. However, previous studies did not show a major advantage of head-mounted devices when compared with handheld devices [40, 5]. Moreover, the current bulkiness of head-mounted devices [32] and both voice control and eye-level hand gestures raise much concern of social acceptability [35, 14]. Therefore, we do not consider smart glasses in this research.

Both VAR and HAR solutions increase the awareness and the connection with surroundings while exploring an urban environment. Nonetheless, both present advantages and constraints. In our study, we aim to answer how these two complementary interfaces can coexist on one system, overcoming the limitations of each.

SYSTEM DESIGN

In this section, we describe the design of the research prototype, comprising VAR and HAR interfaces for urban exploration and a basic mechanism for switching between the two interfaces. We adopted matured technologies at this explorative stage rather than pursuing technical novelty.

VAR Interface

In addition to the common features on most VAR applications (e.g., mapping virtual content to a real scene through camera preview), our VAR interface presents two additional features: 1) precise positioning of the billboards and 2) an aim-and-shoot technique for one-hand operation. Relying on the device location, orientation, and GPS coordinates of the POIs, ordinary VAR applications derive only the relative angles of the surrounding POIs to the current location. It is very common to see a billboard indicating a POI that is invisible from a current location due to long distance or occlusion. To overcome this issue, we incorporated 3D models for checking the visibility of POIs and placing the billboards on relevant locations, such as elevating the billboard of a clock tower (See Fig. 2).

The second feature is an aim-and-shoot technique for enabling one-hand operation, which is preferred by most users in mobile interaction [20]. Ordinary VAR interface design would require one hand for holding the device and the other for clicking the billboards near the center of the display. Our aim-and-shoot interface includes a crosshair icon in the center and a shooting button on the edge of the screen (See Fig. 2). To select a billboard, users align the crosshair with the target billboard and click on the shooting button that is conveniently reachable by the thumb of the device-holding hand.

[2]https://maps.google.com/

[3]https://www.tripadvisor.com/

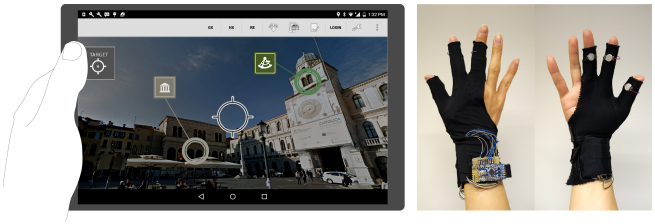

Figure 2. Left: Exact POI locations on the VAR interface are indicated by colored circles which are attached to square billboards featuring POI category icons. The billboards are floating on the screen to avoid obscuring the POIs. The interface features a "TARGET" button on the top-left corner for one-hand selection, and a crosshair icon in the center for aiming at billboards. Right: The SSH Glove. The electronic components were covered with an extra layer of fabric during the experiment.

State	*Activation (ms)*	*ISI (ms)*
Browsing	100	900
Matching	100	400
Multi-Matching	100	100

Table 1. The activation and interstimulus intervals for each tactile cue designed for three states in the interaction.

HAR Interface

The HAR interface has three main components: a hand-worn tracking device for enabling mid-air gestures, a headset for delivering audio information, and a mobile device for location and system logic, etc. The same mobile device also supports the VAR interface. Users do not need to interact with the mobile device when using the HAR interface.

We deployed the SSH Glove (See Fig. 2) for hand tracking. The glove features hand orientation and finger pose sensing, and it gives vibrotactile feedback on three fingers. A 9-axis inertial measurement unit (IMU, InvenSense MPU-9150) is mounted on the back of the hand for sensing pointing orientation. Three flex sensors (Spectra Symbol flex sensor) on three fingers (thumb, index, and middle finger) detect the finger pose. Three vibrotactile actuators (Precision Microdrives 10 mm shaftless vibration motor) provide vibration amplitude at 2.4 g and 250 Hz frequency, which falls in the optimal sensitive range of our tactile perception [17]. The Arduino Pro Mini-based glove is wirelessly connected with the mobile device through Bluetooth.

On the HAR interface, information is encoded and delivered through auditory and tactile features. Auditory features include spoken title and description, as well as auditory icons for indicating the POI category. For example, religious POIs were associated with the sound of a church bell and Gregorian chanting. As for the tactile features, burst rate was the primary source for coding. We implemented three tactile cues for different states of the interaction. The cues have the same activation interval with varied interstimulus intervals (ISIs) (See Table 1). The cue design is in line with previous research [17, 10].

The interaction between the system and the user does not start until a pointing gesture (index finger straightened, middle finger bent, and the hand is horizontal regardless of the arm pose) is performed. When pointing, the user can feel the browsing tactile cue presented on the index finger, indicating the activation of the "browsing" state. If the pointing orientation matches the direction of a POI, a faster matching cue is presented to indicate a "matching" state. Meanwhile, the auditory icon and the spoken title of the matched POI is delivered over the headsets. While in "matching" state, the user can make a selecting gesture (bending index finger when in pointing gesture) for receiving the spoken description. When listening to the description, the user can freely move the hand. Nevertheless, a rejection gesture (only thumb bent) can skip the playback of the spoken content. We followed the recommendation for mid-air pointing gesture design [38]: bending index finger (while pointing) as primary gesture for selection and bending thumb for secondary tasks.

One challenge in designing non-visual HAR interfaces for target acquisition tasks is to identify and indicate overlapping targets, which can be very common in an urban environment. When there are more than two POIs falling into the glove "viewing angle," the multi-matching cue is triggered on the index finger, which suggests the user to approach in the pointing direction to discover the POI hot zone. Meanwhile, the title and auditory icon of the best-matching POI (i.e. with the smallest angular difference in pointing) is delivered, as suggested in [24].

Interaction Design for Switching Interfaces

We implemented a simple switching mechanism requiring minimum actions as a probe to investigate the adjustments and efforts required from the users during interface switching. By detecting the device orientation and screen cover status, we can interpret users' intentions on which interface to use. The design was inspired by the switching between 2D map (screen facing up) and VAR view (back camera facing front) on Nokia's City Lens[4] application.

When the mobile device is in landscape orientation and the screen is not covered, VAR is activated, and HAR is disabled to avoid false positives. To switch from VAR to HAR, the user simply covers the device screen (e.g., by storing the device in a pocket). To return to VAR interface, the user can simply uncover the tablet. This mechanism was reliably done with on-device accelerometers and proximity sensors that are commonly available on modern mobile devices.

A smooth transition was implemented for the interface switching as well. When a POI is activated after selection on HAR, switching to VAR would not interrupt the on-going spoken description while visual media of the POI is available on the screen. This allows users to continue listening to the audio content while examining visual media.

Occlusion Geometry

A 3D model can enable visibility checks on surrounding POIs to avoid the confusion resulting from occlusion. Receiving content regarding a POI while pointing at an irrelevant building that occludes the designated POI can be confusing in both VAR

[4]https://help.here.com/wp8/citylens/

and HAR use scenarios. Furthermore, information regarding distance and elevation of a POI can be displayed on VAR (e.g., the billboard of a clock tower far from the current location can be smaller than nearby POIs and elevated above ground floor).

The 3D model runs in the background on our system and is not meant to be accessible for the users. While the point-and-select technique on both VAR and HAR enables direct interaction with the environment, ordinary touchscreen-based interaction for 3D models remains focused on the screen, thus is excluded from this research.

EVALUATION

Goals

One goal of the study was to check whether aligning geo-located information with users' pointing and viewing direction would result in a better visiting experience than that on an ordinary 2D map interface. Common 2D geo-localized digital maps require users to shift the focus between the surroundings and a touchscreen-based interface, which typically results in a head-down interaction. Thus our sample was split into an experimental group using our research prototype and a control group using the popular Google Maps application. The following hypothesis was formulated:

- (H1) The VAR/HAR interface leads to better visiting experience than the 2D map interface.

A second goal was to examine the efficiency and the perceived advantages of each modality, as well as users' preference for any of them under different tasks. Three different conditions were experienced by the experimental group: using VAR alone, using HAR alone, and using both VAR and HAR. Connected to this goal, we wanted to consider whether the different modalities suited different task types differently. We used location tasks where participants locate a pre-assigned POI that they do not know beforehand, and information tasks where the participants fetch information about a POI that interests them. The former task reproduced a situation in which the participant needs to find a specific POI s/he heard of, while the latter simulates a situation in which the participant wants to know more about an object that captured his/her interest. We then expected that:

- (H2) VAR is more efficient than HAR in location tasks but less efficient in information tasks.

A third goal was to observe the switching between interfaces, specifically the circumstances under which the switching behavior occurred (e.g., whether participants needed to stop to do so).

Setting and apparatus

The research prototype consisted of an 8-inch compact tablet (nVidia SHIELD K1), the SSH Glove, a headset, a pocket for storing the tablet when not in use, and a photo camera. VAR interface was presented on the tablet, while HAR interface was presented on the SSH Glove and via the headset. In the control group, participants were endowed with the tablet running the geo-localized Google Maps application loaded with the same set of POIs as in the experimental group (See Fig. 3).

Figure 3. Left: The apparatus employed for the experiment. a) The shoulder pouch for carrying the mobile device. b) The SSH Glove. c) Headset. d) The mobile device. e) The camera. f) The shoulder pouch for carrying the camera. Right: A participant with all the apparatus.

	Experimental Group			Control Group
	VAR	HAR	VAR/HAR	2D Map
Location Task	1	1	1	3
Information Task	1	1	2	4

Table 2. The study design. The table shows the number of location and information tasks that participants were asked to perform in the different conditions.

Experiment Design

The study had a mixed design. The interface modes (VAR only, HAR only, and both VAR and HAR) were treated as a within-subject variable. The separation of modalities would help us understand what kind of augmentation is helpful to a given activity. The experiment was conducted in three nearby squares that enabled the separation of the three interface modes (See Fig. 4). On the other hand, the presence of augmentation (VAR/HAR vs. 2D map) was treated as a between-subject variable. This is due to the need for contextualizing the assessment of a new prototype by comparing it with the technical solution that represents the current standard. Participants in both groups executed the same total number of tasks (See Table 2).

Data

Visiting Experience. Visiting experience was measured using a post-experience questionnaire that included 18 items to be answered on a 6-point Likert scale (See Table 3). The pleasantness of the visit (i.e., the extent to which the experience was enjoyable, amusing, and fun) was assessed by 4 items (adapted from [22]). The interference or distraction from the system was assessed by 2 items (adapted from [19]). The capability of the interface to build a sense of presence[5] in the augmented environment was assessed by 4 items (adapted from [9]). Finally, seven items aimed at evaluating the usability of the system, namely the comfort of use (3 items), the usefulness (4 items), and its responsiveness (1 item). Participants were also asked to comment their experience by answering four questions in a

[5]Sense of presence refers to the extent to which the user feels directly immersed in the surroundings compared to perceiving a tool mediating experience of the surroundings. We investigated whether the augmented devices improved the sense of immersion compared with the baseline or rather diminished in the short-term type of experience.

Figure 4. The experiment took place in three nearby squares. An additional smaller square on the top right corner was used for training. Predefined POIs are marked in pink labels. The red dots indicate the starting location on each square.

Figure 5. A set of three pictures for recognizing the target POI in the learning test.

Visiting Experience
- Overall, the visit was pleasant
- Overall, the visit was amusing
- Searching for the POIs was enticing
- Using the system was amusing
Connection with the environment
- Using the system distracted me from the surroundings
- Using the system interfered with the main task
Presence
- I had the impression that the information provided by the system were well integrated with the surroundings
- I had the impression that the system built an environment rich of information
- The system did not hamper the observation of the elements in the surroundings
- When I was using the system I had the impression that I was interacting directly with the POI
Usability
Responsiveness - The system promptly responded to my input
Comfort of use - When I was using the system I felt observed
Comfort of use - The system is bulky
Comfort of use - Using the system had limited my movements
Usefulness - The system helped me identifying the PoI around me
Usefulness - I learnt a lot of new things about the landmarks of the city
Usefulness - This system can be a valuable support for tourists
Usefulness - I would like to use this system for other visits in the future

Table 3. The post-experience questionnaire for measuring visiting experience.

semi-structured interview. More specifically, they were asked what impressed them about the experience, what were the strengths and weaknesses of the application, what were the most innovative elements and whether they had the feeling to be more focused on the devices.

The visiting experience was additionally measured using a learning test in which participants were asked to recognize the target POI from three similar pictures (See Fig. 5). Among which, only one was taken from the target POI in the location task. We assume that using VAR/HAR interfaces (focusing on the target landmarks) to explore the environment would result in participants memorizing more details regarding the viewed subjects than using touchscreen-based 2D map (focusing on the touchscreen) [13].

Efficiency. The task completion time was extracted from the synchronized video recordings, while the number of POI selections in each task was derived from system log. Efficiency might not be the most relevant index in the context of urban exploration. However, if the time to complete the same task between two different interface is very different, this may indicate that one of them requires much more effort to use.

Switching. The video recordings were analyzed to identify the occasions and the circumstances (still or on the move) when interface switching took place in the combined VAR/HAR condition.

Procedure

Preparatory phase. The participants were first debriefed on the overall experimental purposes and procedure, signed an informed consent form, and answered a brief questionnaire for background information. Then they were trained to use the devices and practiced both location and information tasks until they were confident in using the interfaces. To remove the bias of participants' prior knowledge about the city, we included in the POI selection several unpopular POIs that are least cited by most tourism apps and websites. The participants were shown the POI list from which they selected for each square a POI that was both unknown to them and seemingly of some interest. The selected POIs (that are unknown to the participant) were to be used as target POIs in the location tasks. The participants were then assigned to a starting condition, either with VAR or HAR, and a starting square (See Fig. 6), both in counterbalanced order.

Experimental phase. The participants were walked to a predefined starting point in the square. They were asked to locate the POI that was selected in the preparatory phase, walk to it, and indicate it to the experimenter (location task). Continuing to the information task, they were asked to explore the available POIs and find information about a specific POI of their choice. When the POI was determined and selected, they again walked to it to indicate it to the experimenter, and finally took a picture of it with the camera provided. After completing both tasks, the participants were asked to answer the post-experience questionnaire and to take the learning test. Then, the participants moved to the second square and repeated the same procedure but with the other interface mode.

	Experimental Group		Control Group					
	M(SD)	*Mdn*	*M(SD)*	*Mdn*	*U*	*z*	*r*	*p*
Visiting Experience	4.78(0.91)	5.00	4.82(1.25)	5.12	78.5	-.558	-0.09	0.58
Connection	4.16(0.78)	4.50	4.95(0.72)	5.00	3.0	-4.22	-0.01	0.021
Presence	4.28(0.93)	4.25	4.30(1.05)	4.37	88.00	-.096	-0.47	0.94
Usability	4.22(0.48)	4.38	4.80(0.65)	5.00	30.00	-2.87	-0.70	0.03

Table 4. Results for the Mann-Whitney test. For the reader's convenience, means, standard deviations, and medians are reported with respect to the dimensions assessed by the post-experience questionnaire for both the experimental and control groups.

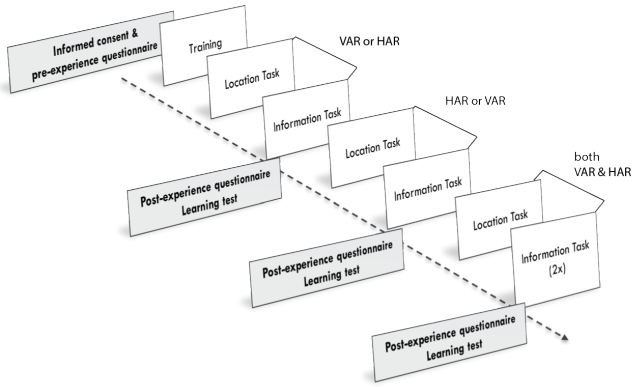

Figure 6. Steps composing the study procedure, with the three stages of the study plus the training.

Finally, the participants moved to a third square where they were free to choose which interface to use as well as switch interfaces during the tasks. When the information task was completed, the participants were forced to switch from the current (favored) interface to the other (less favored) and perform again the information task with a different target. This way, we assured we could observe whether using the less favored interface hampered the performance. At the end, they answered the questionnaire, took the learning test and answered a brief interview. The whole procedure took about 1 hour and 15 minutes per subject.

Participants in the control group performed the same sequence of location and information tasks in the three different squares, always using the same 2D map application.

Participants

A total of 36 volunteers were randomly divided into two groups. The mean age was 25.3 years (*SD* = 2.3) in the experimental group (18 participants, 9 women) and 25.1 years (*SD* = 3.2) in the control group (18 participants, 9 women). All participants reported having never or rarely used any vibrotactile or VAR device and usually relied on guidebooks (30), apps (26), or guided tours (3) when visiting a city. All participants were recruited by word of mouth and through advertising on University pages and social networks. Participants' prior knowledge of the city was not an exclusion criterion, as target POIs were not main city landmarks.

Analysis

The video of participants recorded by a trained experimental assistant and the video of the tablet screen recorded by AZ Screen Recorder[6] were synchronized and embedded into a single video clip. The video analysis followed a top-down approach based on a coding scheme previously agreed upon by three of the authors [33] and was performed by one person who was not in the development team to avoid unconscious bias. Event types were marked by explicitly observable cues. The video clips were analyzed offline using The Observer XT 12 by Noldus to annotate the time during which participants were actively engaged with each task (all conditions), favorite interface mode on the third square (VAR/HAR combined condition), and the episodes when they switched from one interface to the other in VAR/HAR condition. Specifically, the time to complete a task was measured starting from the moment when the participants raised the tablet (when using VAR or 2D map) or the gloved arm (when using HAR) to the moment when they approached the POI and identified it to the experimenter. Walking time was included in the measure because searching for the POI on the site (by the given information) was part of the task. Moreover, participants were also using the system while wandering between the starting point and the target. The switching occurrences were further analyzed in order to identify their circumstances and the modalities that were switched.

The items in the post-experience questionnaire were grouped based on the visiting experience dimension they were supposed to measure (i.e., pleasantness of the visit Cronbach's α = .937, connection with the environment Cronbach's α= .649, presence in the digital information space created by the application Cronbach's α = .72, and usability Cronbach's α = .584); the average score of each dimension was calculated.

Visiting Experience. To compare the visiting experience (H1), a Mann-Whitney test was run, which compared the signed-ranks of the scores at the post-session questionnaire in both groups. Here, the scores refer to the first time participants had experienced each system (i.e., after the first stage for the control group and after the third stage for the experimental group). The scores from the learning test for both groups were also compared. Finally, the ratio between the time spent by participants gazing at the tablet screen and the task duration was calculated and compared with a Mann-Whitney test for VAR and 2D map conditions.

Efficiency. To test the efficiency of the interface modes in the different tasks (H2), the time to complete the location task and the number of POIs selected to complete the task in VAR and HAR conditions were compared with a Wilcoxon test.

[6]https://az-screen-recorder.en.uptodown.com/android

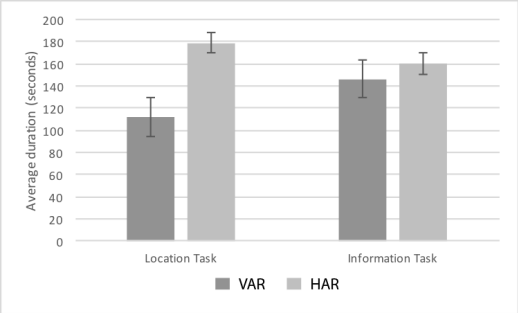

Figure 7. Average time to complete the location and information tasks using VAR and HAR. Bars indicate the standard error.

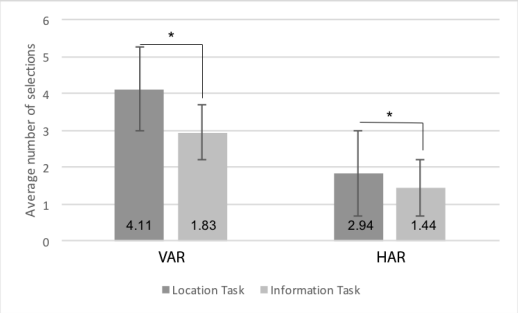

Figure 8. The average number of selections made while performing the location and the information tasks using VAR and HAR. Bars indicate the standard error. *$p < .05$

Switching. To investigate the efforts in switching between VAR and HAR, we observed the switching episodes presented in the video-recordings, especially whether they needed to stop to perform the switching.

RESULTS

Visiting Experience (H1)

The questionnaire scores regarding pleasantness and presence do not seem to differ between the two groups (See Table 4). The only noticeable difference resides in the dimensions of connection and usability. The control group declared they experienced the system as significantly less distracting and interfering, whereas the experimental group found the system easy to use. To see whether the lower scores from the experimental group in these two dimensions were to be attributed to a particular mode, the difference between the three interface modes was analyzed. No significant differences emerged.

As for the learning test, all conditions gave very good results with percentages close to the maximum (respectively, 94% in VAR, 89% in HAR, 94% in combined VAR/HAR, and 89% in the control condition); hence, no further statistical analysis was run.

Efficiency (H2)

Of the two efficiency measures, task completion time (See Fig. 7) and the number of selections made (See Fig. 8), differences in VAR and HAR were found only in the latter. Contrary to the hypothesis, for the location task, a higher number of selections was found in VAR (M = 4.11, SD = 3.02, Mdn = 3.5) than in HAR condition (M = 1.83, SD = 1.15, Mdn = 2), Z = -2.579, p = .010. Conforming to H2, for the information task, fewer POI selections were made when using HAR (M = 1.44, SD = 1.04, Mdn = 1) than VAR (M = 2.94, SD = 2.5, Mdn = 2), Z= -2.417, p = .016. In synthesis, regardless of the task, using VAR led to more POI selection than using HAR but with similar task completion time.

Switching

When using both VAR and HAR combined, the participants spontaneously switched interfaces more frequently from HAR to VAR (16) than the contrary (9). Switching interfaces required participants to stand still in the majority of the cases (20 out of 25). No difference was found between the task completion time when using the preferred (M = 136.18 seconds SD = 34.58) and less preferred (M = 124.11 seconds SD = 27.20) interface, suggesting that participants could effectively use both modes despite their preferences.

Participants were inquired in the final interview about the reasons why they spontaneously switched interface. Four of the participants switching from VAR to HAR reported that they preferred to listen to the audio rather than reading the textual description of the POIs (3):

> *"This coupling was ideal (VAR+HAR). You can scan the points of interest with the tablet and then listen (to the description) with the headsets, so you can look directly at the point of interest [P11]"*

> *"If possible I would have used only the glove (HAR) ... but since I arrived in a place with many things I didn't know, I wanted to get an overview first ... at that point it is less intrusive to point and listen instead of reading [P33]"*

On the other hand, participants who switched from HAR to VAR mentioned that they wanted to verify the directions provided by HAR (4), to better localize the POI (4), to get information faster (1), to be more in control of the application (1) or because they found VAR more intuitive (1). For example:

> *"(With HAR) my hands were free. I could move more easily ... (I switched to VAR because) I wanted to have a confirmation that I had found the right monument [P31]"*

> *"I like the glove better. I find it easier to use and less bulky ... At some point I could not find a point of interest, but with the tablet (by switching to VAR) I managed [P47]"*

When asked about the strengths of each AR mode, participants seem to find that VAR enables the syncretic overview of all POIs available in the surroundings (3), with an accurate localization of the landmarks (2), while HAR is considered more amusing (5), innovative (3), and able to get information without a screen interposed between them and the environment, enabling a head-up interaction (7):

> *"(The glove) is less bulky... You don't even feel like you are using a tool and even the vibration was pleasant... It felt as if it was an extension of your body [P16]"*

> *"(With the glove) you can interact directly with the environment... you get a concise description of what you see.*

Sometimes you visit places and see things and wonder what they are. Here you only need to make a simple pointing gesture and you get to know the basic info [P27]"

"You can use it (VAR) by just turning around and see in real-time all the points of interest [P52]"

"The tablet is good to see where are the points of interest. With the glove if there are more hidden or small points of interests is difficult to find them. I think the glove is better to find bigger monuments and landmarks ... if used together (with VAR) is helpful [P47]"

Other qualitative remarks

To better understand how the system affected the general visiting experience it is interesting to look at some more qualitative remarks. In particular we asked about perceived weaknesses of the VAR+HAR interface. Here participants mainly complained about the discomfort caused by the too many devices:

"The bulkiness. Having to carry the tablet and the glove with this warm weather... [P28]"

"You have to handle many things and your hands are always busy [P44]"

Additionally some participants reported to be very focused on the technology and that caused some distraction:

"At some point I was very focused on the glove ... I was focused on where my finger was pointing ... I was thinking at how to use my glove, I mean, my hand, as a pointer... but after a while it became comfortable and natural [P27]"

Other times the system was found difficult to use due to the noise of the external environment:

"Sometimes the noise from the crowd didn't allow me to hear the audio descriptions [P32]"

In general, however, the system was well accepted:

"I feel this system is very innovative. You can use an object absolutely not bulky as a glove and another very common as a tablet to get the maximum of AR ... I mean I don't have to go around and look tons of things on my guide, but I can use this special audio-guide (HAR) to see what's around me and only use the tablet to go to a specific place. [P52]"

"Even though physically bulky (VAR+HAR), you are still aware of what's going on around you, at safety level also ... You can easily notice other things (from the environment) while receiving the suggestions of the interface [P44]"

DISCUSSION

VAR and HAR interfaces have complementary strengths in helping users interact with information while maintaining focus on the environment. In this study, we explored the effect of providing users with an urban explorer interwoven with both VAR and HAR. We discussed how participants develop their strategies in using these interfaces in response to the immediate situation as well as providing insights into how each interface could be improved. The observed spontaneous switching confirmed the usefulness for a multi-interface system in a dynamic environment.

Visiting Experience

In terms of visiting experience (H1), the multi-interface system did not differ from the more popular 2D map system even though the former was perceived as more distracting and more difficult to operate. This "distractibility" was probably due to both the number of devices to be managed and the novelty of the system. The lower usability scores of the multi-interface system are consistent with the sensation of being focused on the system itself. However, it is interesting that the other scores were not affected, and the system was considered pleasant and allowed a sense of presence in the information space.

Results from the learning test showed that both groups had high accuracy, indicating that the distinguishability between the pictures was too high. Even though no further insights were gained, we reported this test for the record and the community may find this method useful.

Efficiency of different modalities

Regarding efficiency, H2 was partly confirmed: the different interface modes were indeed used differently, although this did not depend on the task. Basically, VAR always led to selecting more POIs than HAR. Although this did not lead to longer task completion time. While filtering for wanted information through HAR requires some time to listen to the audio information for each selection, skimming for information on VAR, due to the high bandwidth in processing textual information visually, could be quickly done by repeating the following: POI selection, information skimming, quitting, and selecting the next. Timewise, the cost of selecting a wrong POI on VAR was lower than in HAR. So in VAR the time wasted in opening wrong PoIs was eventually compensated by the higher speed at which they could be found compared with HAR. Hence the higher number of selections did not result in longer task completion time.

Switching between modalities

Spontaneous interface switching during the exploration revealed the usefulness of providing complementary interfaces on one system. Although interface switching causes participants to stop, sacrificing both time and physical effort, the majority of the participants spontaneously switched an interface when they felt the other interface was more suitable for a concurrent situation. The situation, be it subject feeling or environmental factors, could be so dynamic that it drove participants to switch to the other interface in the next moment. Equally important to be considered in the decision making is the effort required from the interface switching mechanism. This observation shows that the proposed switching mechanism did not present a high cost which could stop participants from switching interfaces. Moreover, the result corresponds to previous research on users' behavior in urban exploration, where their needs are dynamic and evolving, and the tools should support switching between interfaces [37]. As praised by participants, these two interfaces are complementary and it was easier to locate POIs with them working together.

Effects of Interaction on Strategy Development

The fundamental differences between the modalities have an impact on participants' decision making in choosing which interface to use. People seem to be more comfortable and confident in using visual interfaces than non-visual ones. Some of the participants switched from HAR to VAR because they wanted to check the direction given by VAR, to better localize the POIs, or to be more in control of the application. However, none of these was the reason why people switched from VAR to HAR although direction misalignment due to sensor error is pervasive. This could result from the fact that people are more familiar with and confident in using visual interface, hence more comfortable with certain errors. On the other hand, HAR did provide unique experience in the exploration that a VAR cannot achieve. Some participants developed a strategy that they localized a POI through VAR but switched to HAR because they wanted to listen to, instead of read, the description of the POI while visiting it. Most of all, although both VAR and HAR enabled head-up interaction, HAR was preferred for it did not interpose a screen between the participants and the surroundings.

The ratio between the number of selections made and the total time also suggests something that is aligned with the above observation: VAR is mainly employed for fast access to information, while HAR serves purposeful exploration of a given POI. While using VAR, participants tended to select multiple POIs for fast access to information. Although the title of the POIs was visible on the billboard, participants skipped reading the title. Instead, they simply selected POIs one after another and examined each POI's detailed view on the pop-up window until the correct one was found. This could result from the fact that the positions of POI billboards (where the title was) reacted to the tablet movement, while on the pop-up window, the title was always displayed on the same position. Participants might have preferred monitoring the same position for POI title, although with extra clicks, than tracing the moving titles. On the other hand, the selections made with the HAR interface seemed to be more purposeful, as participants tended to pay attention to the spoken title and select only when the title matches the target. Participants confirmed this strategy when they were interviewed, reporting that VAR served as a probe of the indications provided by HAR. In addition, HAR mode seems to be less suitable for supporting the search of a specific POI. Even though the difference is not statistically significant, the time required to complete a location task with HAR is higher than using VAR and with both.

Overall, VAR was chosen for faster access to information because it was perceived more precise and could display multiple POIs at once. HAR was chosen for a more purposeful exploration and experience because it did not interpose a screen between the user and the surroundings.

Limitations

In contrast to our expectations, users reported that the multi-interface application demanded more attention as compared to the 2D map application. Despite the scores recorded by the post-experience questionnaire for the experimental group being overall positive, they were indeed lower as compared to the control group. It should be considered that there were more devices to be handled in the experimental group than in the control group, which also explains why the multi-interface system was perceived as more bulky. Moreover, while most people were familiar with the 2D map application, all participants were novice users of the multi-interface system. Interestingly, participants' subjective reports are in contrast with objective observations: VAR users did not spend more time looking at the screen than 2D map users. Nevertheless, the behavioral difference was obvious: VAR users continued a head-up interaction, whereas baseline users used the tablet with heads down.

Some device-specific limitations on the interaction might affect the overall experience in urban exploration. For example, the glove could be made with more flexible fabric to be more comfortable and robust. Using wireless headsets would also reduce the physical constraints in mobility, especially when there are more than one device to manage. Choosing a more compact mobile device could also improve the mobility. However, a smaller screen limits the amount of information that could be displayed at the same time and the comfort of reading.

Both the hand gestural interaction and the switching of interface could suffer from false positives/negatives that are common on sensor-based interaction. Although we have designed the horizontal pointing gesture as the initiation of the interaction, false positives are still possible on various occasions. Similarly, the mobile device may be held in different poses under various conditions. More careful design is needed for interpreting users' intentions effectively through the sensors as well as the activation and closure of the interaction.

CONCLUSION

We investigated how VAR and HAR interfaces could coexist on one urban explorer system to allow easy adaption to a highly dynamic environment. A research prototype equipped with both interfaces was implemented for studying people's preferences in using either interface under different circumstances, how people manage multiple interfaces, and how the switching of interfaces can be designed. We found that VAR was preferred when faster access to information was desired, while HAR was preferred for a more purposeful exploration. Moreover, HAR enabled users to get information without a screen interposed between them and the environment. We also observed spontaneous switching of interfaces from most participants, which indicates that the availability of both interfaces on one system assisted users to react to environmental change or personal need. The research suggests that a versatile, multi-interface urban explorer design is preferable, which enables users to easily adapt to the concurrent environment by selecting the more feasible interface while maintaining focus on the surroundings.

ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n°601139 and FCT/MCTES LARSyS (UID/EEA/50009/2013 (2015-2017)).

REFERENCES

1. Anupriya Ankolekar, Thomas Sandholm, and Louis Yu. 2013. Play it by ear: a case for serendipitous discovery of places with musicons. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, Paris, France, 2959–2968. DOI: http://dx.doi.org/10.1145/2470654.2481411

2. Matthias Baldauf, Peter Frohlich, Kathrin Masuch, and Thomas Grechenig. 2011. Comparing Viewing and Filtering Techniques for Mobile Urban Exploration. *Journal of Location Based Services* 5, 1 (March 2011), 38–57. DOI: http://dx.doi.org/10.1080/17489725.2010.541161

3. Michel Beaudouin-Lafon. 2004. Designing Interaction, Not Interfaces. In *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI '04)*. ACM, New York, NY, USA, 15–22. DOI: http://dx.doi.org/10.1145/989863.989865

4. Mark Billinghurst and Bruce H. Thomas. 2011. *Mobile Collaborative Augmented Reality*. Springer New York, New York, NY, 1–19. DOI: http://dx.doi.org/10.1007/978-1-4419-9845-3_1

5. Anne Braun and Rod McCall. 2010. User Study for Mobile Mixed Reality Devices. In *Joint Virtual Reality Conference of EGVE - EuroVR - VEC*. The Eurographics Association. DOI: http://dx.doi.org/10.2312/EGVE/JVRC10/089-092

6. Jaewoo Chung, Francesco Pagnini, and Ellen Langer. 2016. Mindful navigation for pedestrians: Improving engagement with augmented reality. *Technology in Society* 45 (2016), 29–33. DOI: http://dx.doi.org/10.1016/j.techsoc.2016.02.006

7. Andrew Dillon, John Richardson, and Cliff McKnight. 1990. The effects of display size and text splitting on reading lengthy text from screen. *Behaviour & Information Technology* 9, 3 (1990), 215–227.

8. Peter Fröhlich, Antti Oulasvirta, Matthias Baldauf, and Antti Nurminen. 2011. On the move, wirelessly connected to the world. *Commun. ACM* 54, 1 (jan 2011), 132–138. DOI: http://dx.doi.org/10.1145/1866739.1866766

9. Maribeth Gandy, Richard Catrambone, Blair MacIntyre, Chris Alvarez, Elsa Eiriksdottir, Matthew Hilimire, Brian Davidson, and Anne Collins McLaughlin. 2010. Experiences with an AR evaluation test bed: Presence, performance, and physiological measurement. In *Mixed and Augmented Reality (ISMAR), 2010 9th IEEE International Symposium on*. IEEE, 127–136.

10. Frank A. Geldard. 1957. Adventures in tactile literacy. *American Psychologist* 12, 3 (1957), 115–124. DOI: http://dx.doi.org/10.1037/h0040416

11. Shigeru Haga, Ayaka Sano, Yuri Sekine, Hideka Sato, Saki Yamaguchi, and Kosuke Masuda. 2015. Effects of using a Smart Phone on Pedestrians' Attention and Walking. *Procedia Manufacturing* 3 (2015), 2574–2580. DOI:http://dx.doi.org/10.1016/j.promfg.2015.07.564

12. Faizan Haque, Mathieu Nancel, and Daniel Vogel. 2015. Myopoint: Pointing and Clicking Using Forearm Mounted Electromyography and Inertial Motion Sensors. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*. ACM Press, Seoul, Republic of Korea, 3653–3656. DOI: http://dx.doi.org/10.1145/2702123.2702133

13. Linda A Henkel. 2014. Point-and-shoot memories: the influence of taking photos on memory for a museum tour. *Psychological science* 25, 2 (feb 2014), 396–402. DOI: http://dx.doi.org/10.1177/0956797613504438

14. Yi-Ta Hsieh, Antti Jylhä, Valeria Orso, Luciano Gamberini, and, Giulio Jacucci. 2016. Designing a Willing-to-Use-in-Public Hand Gestural Interaction Technique for Smart Glasses. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*. ACM Press, New York, New York, USA, 4203–4215. DOI: http://dx.doi.org/10.1145/2858036.2858436

15. Ira E Hyman, S Matthew Boss, Breanne M Wise, Kira E McKenzie, and Jenna M Caggiano. 2010. Did you see the unicycling clown? Inattentional blindness while walking and talking on a cell phone. *Applied Cognitive Psychology* 24, 5 (oct 2010), 597–607. DOI: http://dx.doi.org/10.1002/acp.1638

16. Lei Jing, Zixue Cheng, Yinghui Zhou, Junbo Wang, and Tongjun Huang. 2013. Magic Ring: a self-contained gesture input device on finger. In *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia - MUM '13*. ACM Press, New York, New York, USA, 1–4. DOI: http://dx.doi.org/10.1145/2541831.2541875

17. Lynette a Jones and Nadine B Sarter. 2008. Tactile displays: guidance for their design and application. *Human factors* 50, 1 (2008), 90–111. DOI: http://dx.doi.org/10.1518/001872008X250638

18. Matt Jones, George Buchanan, and Harold Thimbleby. 2003. Improving web search on small screen devices. *Interacting with Computers* 15, 4 (2003), 479–495.

19. Antti Jylhä, Yi-Ta Hsieh, Valeria Orso, Salvatore Andolina, Luciano Gamberini, and Giulio Jacucci. 2015. A Wearable Multimodal Interface for Exploring Urban Points of Interest. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*. ACM Press, New York, New York, USA. 175–182 DOI: http://dx.doi.org/10.1145/2818346.2820763

20. Amy K Karlson, Benjamin B Bederson, and Jose L Contreras-Vidal. 2008. Understanding one-handed use of mobile devices. In *Handbook of research on user interface design and evaluation for mobile technology*. IGI Global, 86–101.

21. Andreas Komninos, Peter Barrie, Vassilios Stefanis, and Athanasios Plessas. 2012. Urban exploration using audio

scents. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services - MobileHCI '12*. ACM, 349–358. DOI:http://dx.doi.org/10.1145/2371574.2371629

22. Sang-su Lee, Sohyun Kim, and Kun-pyo Lee. 2010a. How Users Manipulate Deformable Displays as Input Devices. *Methodology* (2010), 1647–1656.
23. W. Lee, Y. Park, V. Lepetit, and W. Woo. 2010b. Point-and-shoot for ubiquitous tagging on mobile phones. In *2010 IEEE International Symposium on Mixed and Augmented Reality*. 57–64. DOI: http://dx.doi.org/10.1109/ISMAR.2010.5643551
24. Charlotte Magnusson, Kirsten Rassmus-Gröhn, and Delphine Szymczak. 2010. Scanning angles for directional pointing. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services - MobileHCI '10*. ACM Press, Lisbon, Portugal, 399–400. DOI: http://dx.doi.org/10.1145/1851600.1851684
25. Charlotte Magnusson, Kirsten Rassmus-Gröhn, and Delphine Szymczak. 2014. Exploring history: a mobile inclusive virtual tourist guide. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction Fun, Fast, Foundational - NordiCHI '14*. ACM Press, New York, New York, USA, 69–78. DOI: http://dx.doi.org/10.1145/2639189.2639245
26. David McGookin, Stephen Brewster, and Pablo Priego. 2009. Audio bubbles: Employing non-speech audio to support tourist wayfinding. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 5763 LNCS (2009), 41–50. DOI: http://dx.doi.org/10.1007/978-3-642-04076-4_5
27. Jack Nasar, Peter Hecht, and Richard Wener. 2008. Mobile telephones, distracted attention, and pedestrian safety. *Accident; analysis and prevention* 40, 1, 69–75. DOI:http://dx.doi.org/10.1016/j.aap.2007.04.005
28. Rui Nóbrega, Diogo Cabral, Giulio Jacucci, and António Coelho. 2015. NARI: Natural Augmented Reality Interface - Interaction Challenges for AR Applications. In *Proceedings of the 10th International Conference on Computer Graphics Theory and Applications (VISIGRAPP 2015)*. SCITEPRESS, 504–510. DOI: http://dx.doi.org/10.5220/0005360305040510
29. Eamonn O'Neill. 2014. Haptic and audio displays for augmented reality tourism applications. In *2014 IEEE Haptics Symposium (HAPTICS)*. IEEE, Houston, TX, USA, 485–488. DOI: http://dx.doi.org/10.1109/HAPTICS.2014.6775503
30. Martin Pielot, Niels Henze, and Susanne Boll. 2009. Supporting Map-based Wayfinding with Tactile Cues. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '09)*. ACM, New York, NY, USA, Article 23, 10 pages. DOI: http://dx.doi.org/10.1145/1613858.1613888
31. Martin Pielot, Wilko Heuten, Stephan Zerhusen, and Susanne Boll. 2012. Dude, where's my car?: in-situ evaluation of a tactile car finder. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction Making Sense Through Design - NordiCHI '12*. ACM Press, Copenhagen, Denmark, 166–169. DOI: http://dx.doi.org/10.1145/2399016.2399042
32. Dieter Schmalstieg and Tobias Höllerer. 2015. *Augmented Reality: Principles and Practice*. Addison Wesley Professional.
33. Julia Snell. 2011. Interrogating video data: Systematic quantitative analysis versus micro-ethnographic analysis. 14 (05 2011), 253–258.
34. Mayuree Srikulwong and Eamonn O'Neill. 2011. A comparative study of tactile representation techniques for landmarks on a wearable device. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI '11*. ACM, Vancouver, BC, Canada, 2029–2038. DOI: http://dx.doi.org/10.1145/1978942.1979236
35. Ying-Chao Tung, Chun-Yen Hsu, Han-Yu Wang, Silvia Chyou, Jhe-Wei Lin, Pei-Jung Wu, Andries Valstar, and Mike Y Chen. 2015. User-Defined Game Input for Smart Glasses in Public Space. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*. ACM Press, 3327–3336. DOI:http://dx.doi.org/10.1145/2702123.2702214
36. Matthew Turk. 2014. Review Article: Multimodal Interaction: A Review. *Pattern Recogn. Lett.* 36 (Jan. 2014), 189–195. DOI: http://dx.doi.org/10.1016/j.patrec.2013.07.003
37. Tuomas Vaittinen and David McGookin. 2016. Phases of Urban Tourists' Exploratory Navigation: A Field Study. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems (DIS '16)*. ACM, New York, NY, USA, 1111–1122. DOI: http://dx.doi.org/10.1145/2901790.2901795
38. Florian van de Camp, Alexander Schick, and Rainer Stiefelhagen. 2013. How to click in mid-air. In *International Conference on Distributed, Ambient, and Pervasive Interactions*. Springer, 78–86.
39. Yolanda Vazquez-Alvarez, Ian Oakley, and Stephen A Brewster. 2012. Auditory display design for exploration in mobile audio-augmented reality. *Personal and Ubiquitous Computing* 16, 8 (sep 2012), 987–999. DOI: http://dx.doi.org/10.1007/s00779-011-0459-0
40. J. Wither, S. DiVerdi, and T. Hollerer. 2007. Evaluating Display Types for AR Selection and Annotation. In *Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on*. 95–98. DOI: http://dx.doi.org/10.1109/ISMAR.2007.4538832
41. Z Yovcheva, Dimitrios Buhalis, and C Gatzidis. 2012. Smartphone Augmented Reality Applications for Tourism. *e-Review of Tourism Research (eRTR)* 10, 2 (2012), 63–66.