Augmented Paper Maps: Design of POI Markers and Effects on Group Navigation

Jeries Besharat[a,c,*], Andreas Komninos[b], George Papadimitriou[a], Eirini Lagiou[a] and John Garofalakis[a,c]
[a] *University of Patras*
[b] *University of Strathclyde, Glasgow, United Kingdom*
[c] *RACTI, Rio, Patras, Greece*

Abstract. One popular and wide use of augmented-reality based application, is the projection of points of interests on top of the phones' camera view. In this paper we discuss the implementation of an AR application that acts as a magic lens over printed maps, overlaying POIs and routes. This method expands the information space available to members of groups during navigation, partially mitigating the issue of several group members trying to share a small screen device. We examine two aspects critical to the use of augmented paper maps: (a) Appropriate visualization of POIs to facilitate selection and (b) augmentation of paper maps with route instructions for use in group situations. In this paper, we evaluate POI visualization in a lab setting and augmented paper map navigation with groups of real tourists in a preliminary field trial. Our work complements existing literature introducing self-reporting questionnaires to measure affective state and user experience during navigation.

Keywords: Augmented reality • group navigation • mobile maps • tourism • augmented maps.

1. Introduction

Many location based mobile applications today allow users to discover the location of Points of Interest (POIs) in a city. Teevan et al. [24] found that the most common reason of searching for a POI on a mobile device was to get directions to that POI. Typically, map exploration and navigation with mobile devices is done using the small screen space available. While navigation applications for single users are well developed and manage to empower users sufficiently, in many situations users do not navigate alone, but as part of a group. In such cases, the convergence of multiple users over a single small-screen device is problematic, as the information display area is too small to be viewed by all members of the group. Hence, collaborative navigation, where multiple users can offer their interpretation of instructions or make decisions on routes to take, is difficult when using mobile devices.

In this paper we describe our work into two crucial aspects of navigation with an augmented paper map. Our prototype, which is called HoloPlane, shows POIs collected from the Foursquare API. HoloPlane uses real-time and historical data from social networking services (as in, e.g., [12]), to display these POIs in a manner that allows users to understand their popularity under the current temporal context. Users can see their own location on the map as a virtual marker and can select POIs to navigate to. Routes are displayed as a set of virtual lines, aligned with the street structure on the printed map.

In terms of evaluation, first, we present the results of a lab-based evaluation of POI visualization techniques in order to allow users to quickly detect and select recommended POIs to navigate to. In this context, we use the popularity of venues fetched from the FourSquare API (i.e. the number of total check-ins at a venue) as a measure of popularity and assess the effect of POI visualization techniques (3D vs. 2D markers, variable or fixed marker size and marker colouring) on the user's ability to discern between popular and unpopular venues. As a second step, we evaluate an alternative approach for group navigation,

* Corresponding author. E-mail: besarat@ceid.upatras.gr.

based on the augmentation of a physical paper map, during a field experiment with real tourists. Given the lack of literature in navigation of groups, instead of individual users, we focus on the collaborative aspects of group navigation using augmented paper maps and discuss the use of sharable augmented paper maps during navigation, compared to a traditional navigation tool (Google Maps). For the evaluation of our prototype during navigation with groups, we introduce an alternative evaluation approach based on user experience using validated questionnaires. This approach to evaluation has not been employed in related research in the past and we discuss our experience of using this method.

2. Related Work

2.1. Visualisations of POIs and Landmarks in mobile maps

In the 50 years of Geographical Information Systems, digital and mobile maps have traditionally used icons as markers to represent POIs. It is not clear who "invented" the use of markers on digital maps, however their use became widespread with the rapid adoption of GIS software and later on-line cartography services such as Google Maps. Since then, markers in mobile and digital maps have been commonly designed empirically, as both our own and previous surveys of literature [8] yielded very few results in the process of designing POI markers. Chittaro [4] introduced the concept of dynamically drawn POI markers that incorporate contextual information on the represented POI, in terms of the degree in which POIs fulfill filtering criteria. This was accomplished by drawing a green bar on the side of each icon in a 2D mobile map, whose height represented the degree in which a POI matched filtering criteria. Elias & Paelke [8] highlight a lack of literature in POI marker design, having found an extremely small body of literature in this area. Their work examined a variety of landmark marker design approaches that adopt various levels of abstraction (from photos to iconology and words) and propose design guidelines for marker visualisation, in which icons, symbols and words for depicting landmark types is found to be the best approach. The use of photographs of landmarks is recommended as appropriate for representing visual aspect, a finding supported by Hile et al. [10] and also Delikostidis et al. [6]. A significant issue with markers on digital maps, particularly affecting mobile maps due to the small screen limitation, is the presence of high volumes of markers in the map view. Several approaches for limiting the clutter on maps have been proposed, most of which focus on the heavy context-aware filtering of visualised POIs, in order to reduce the displayed volume. A visual approach is to cluster markers and represent these with an aggregate marker symbol. These approaches are reviewed in [11].

2.2. Augmented Reality maps and Navigation

Traditionally, paper maps have played major roles in conveying spatial information and guiding people around in space. However, this standard experience could be enhanced and improved, as it is shown that augmented paper maps could be used to develop interactive paper maps that will provide added values services for tourists [18]. Stroila et al. [23] demonstrated an AR navigation application, which allows users to interact with transit maps in public transit locations and vehicles. They created a system where users could use their device and scan a subway transit map and the application would highlight the current as well as already past stops during a journey. The authors however did not evaluate their system with actual users.

In [17] the acceptance and usability of an AR system that provides pedestrian navigation through a combination of mobile devices and public displays are studied, but with focus on single users and not collaborative use. A lack of research on collaborative use is apparent in most literature on this subject. The effectiveness of navigating to POIs with an AR browser and a 2D digital map interface is studied in [7]. It is found that although the use of AR with a digital map did not offer any advantages to performance, users preferred this mode strongly as it doesn't lock users into one type of interaction (i.e. just using mobiles, or just using maps).

Other research has identified a range of issues concerning the use of AR and magic lens interaction (modification of the image displayed on the screen by applying modifications in real time through the viewing region). One is the dual-view problem in magic lens viewing [5], where users have to shift their attention between the mobile screen (magic lens) and the background augmented object. This causes difficulty in matching the mobile view with the background, as the mobile view appears at a different zoom level than the background object, hence posing cognitive difficulties to the user. A further issue arising from natural use of the mobile device, is the

angular difference in the user's view of the background object and the device (e.g., the background object might be perpendicular to the ground as in the case of a fixed poster, while the mobile screen might be tilted in varying degrees, for example when the user holds the device up high to bring a tall part of the background object into view). A further issue concerns the size of the augmented object, in this case the background map.

In [9] it was found that static peephole interfaces for maps are better than magic lenses, when the area of the map to be explored is small. As the size of the map increases, the differences even out and in fact, the magic lens interface becomes better to use in larger maps. The researchers obtained their findings using physical map sizes that are considerably larger than the typical handheld map (the smallest map used was 1.38 m × 0.76 m), making these findings applicable to large maps, of the kind that would be placed on a wall as a poster, or on a public display.

Finally, in [23], researchers find that item density can have an effect on how much time users spend looking at the background object, compared to the magic lens view. It was found that for low item density situations, users tended to focus more on the background object, confirming a previous experiment [21] where users focused more on the magic lens view, above a certain item density threshold.

2.3. Group Navigation with Mobile Maps

In [3] the problems faced by tourists during holidays are outlined. The most common problems in an unfamiliar place, are what to do and when. The researchers explore how tourists solve their problems by relying on sharing the visit with other tourists (79% of leisure visits involve groups of two or more) and how they worked as a group by using digital technologies. The leisure activity seemed to be less important than the fact the tourists spent significant time with others. As a result, technologies that are woven into this sociality are likely to be used in preference to those that are not.

Reilly et al. [20] examined how groups of two share a single device during a collaborative indoor way finding activity. They developed two basic interfaces (one that combines map and textual descriptions and a textual interface that numbers the route description) in order to conduct the experiment. Their analysis on the results showed that the application's interface impacts the strategy users followed to complete the tasks. They found that some pairs heavily favored specific navigation strategies or sharing styles. This emphasizes the importance of group dynamic on the use of spatial applications.

A set of requirements for mobile indoor navigation systems that support collaborative path finding tasks is presented in [2]. The researchers observed and analyzed the actions participants performed such as walking, pointing, looking etc. and found that the pointing action, as a communication purpose, occurs much more in groups. Furthermore, the number of people involved in a group does complicate the process of completing the task. 76.4% of the participants stated that positioning and navigation signs helped them to find their target locations. There is very little relevant literature that discusses group navigation aspects using AR.

In [16], researchers augmented a map with POIs (but not navigation instructions) using a device as a magic lens as part of a pervasive game. They found that augmented maps offer advantages to groups as a collaboration tool, since groups that used them found it easier to establish common ground than groups of users who used only a digital map. Further work in [15] included use of multiple devices on the same map, which found that up to two devices are usable without causing issues. It was also shown that the ability to cluster and collaborate over the physical map enhanced the "feelgood" factor between group members. Neither [15] nor [16] seem to consider the dual-view problem, item density or map size for their effect on the usability of the AR maps. A significant shortfall of studies like [15] and [16] lies in the fact that only qualitative data was obtained by the researchers, in the form of interviews, coupled with their own direct observation. As pointed out in [1], these methods suffer from potential subjectivity bias and also from the researchers' own bias. This observation is highlighted again in [19], where it is found that the user's own context (i.e. whether they see themselves as a future user of the system under evaluation) can place a strong influence on the reported assessment of a system's usability. Hence, the findings in [15] and [16] provide a good insight into the usability of AR maps for collaborative use, but have to be considered as incomplete.

As can be seen, the use of AR maps is an on-going subject of research with many unanswered questions, in both the cases of single users and collaborative use. Our main focus, for this paper, is placed in two concerns: First, to add to the small body of literature on visual marker design, this time in the context of AR maps, a topic that has not been addressed elsewhere. Second, to add to the small body of literature in user

experience during AR-assisted group navigation, taking a different approach to evaluation (i.e. using self-reporting questionnaires that have been validated for effectiveness, to assess affective state and user experience). This element is entirely missing from existing literature, where results are largely based on direct observation and interviews.

3. The HoloPlane Prototype

Our prototype (HoloPlane) is built using the Qualcomm Vuforia SDK and Unity for Android. To present POIs to users, we leverage from our previous work on a platform for aggregating social network data [12]. Users can log in to a dedicated web site using their home PC, where they are able to select an area of interest using a map interface. For the area selected by the user, we obtain a printable map raster image (by querying the static maps procedure of the Google Maps API), which is also used as a recognition target by the mobile client. For this, the map image is sent to the Vuforia Cloud Recognition service and our database keeps a record of the recognition targets created by the user under her account.

For the area selected by the user, we collect and store historical information from FourSquare using the approach described in [12]. Automated queries to the FourSquare [1] API for venues within the geographic bounds of the user-selected map are fired every 30 minutes and the data is saved in the HoloPlane database. This process allows our system to query the collected data under temporal context, in order to present venues based on their popularity for the given temporal context (e.g., "show popular venues for Tuesday 3pm in this area"). Once the user has selected one or more areas that interest her, generating her personal "map collection", she can then print out the maps and use them with her mobile client. Alternatively, the user can save maps (and related POIs) that other users have created into their personal collection on the website. The client requires the user to log-in and is hence able to download the appropriate image recognition target fingerprints that this user has specified for her maps, as well as all the related POI data from our database. The system architecture is depicted in Fig. 1.

The printed map does not require special markings to be recognized by the device, as it is a recognition target in itself. When the application detects the map, it connects to our server and fetches the required POI information. This is overlaid on the map image along with a marker that shows the user's location. Users can select POIs and get further information from our database for each one. The main interface of the application consists of five buttons that are placed on the top area of the screen and one informative panel on the bottom area of the screen. With this layout, we developed a service that conveys a range of contextual information to the user in a multi-layered view.

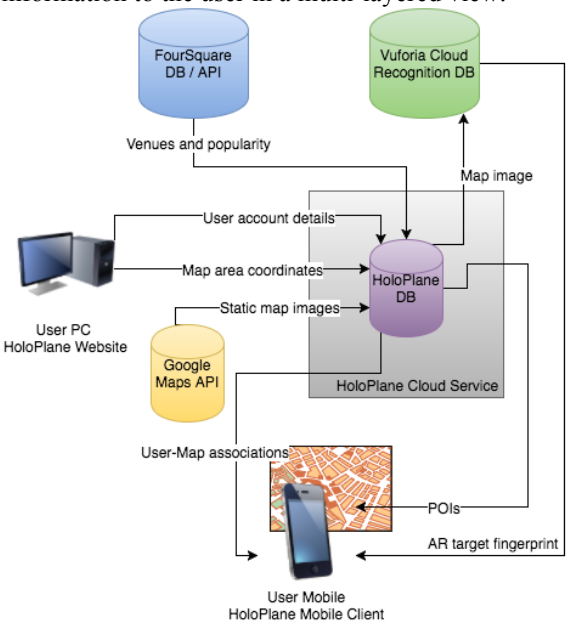


Fig. 1. The HoloPlane System Architecture

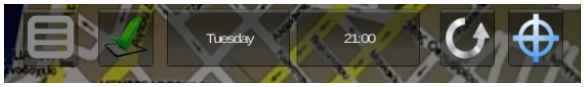
The graphical elements in brackets are shown in Fig. 2 (top). Layer 1: This layer is responsible for overlaying the POI information retrieved from our server. The POIs are presented as 2D or 3D markers that indicate the category the POI belongs to (1). The markers can be colour-coded, or vary in size, to indicate whether they are popular or not, depending on the current time and day. The navigation route (if selected) is also shown (2), using virtual lines, aligned with the street structure on the map. The user's position, which is determined using the devices' GPS sensor, is also displayed as an arrow (3). Layer 2: This layer has all the UI control and is split in two sections, the top UI control bar (Fig. 1 middle) and the bottom navigation panel (Fig. 1 bottom). In the UI control bar, button (4) shows a list view with the names of all the POIs currently on the device screen. Users can select a POI from that list to identify it on the map. The application then "scales up" the POI marker, to help

[1] FourSquare: http://www.foursquare.com

the user identify it. This helps users to find POIs by name and to select POIs in cases when they appear too small (device far from the map) or when many POIs are clustered together. Button (5) shows a popup panel, which allows the user to filter the POIs by category. Buttons (6) and (7) allow control over the temporal context colour-coding, by allowing users to display popularity information for specified days of the week and times (hours) of the day, selectable through drop down lists. Button (8) refreshes the information each time a user selects different values from the drop down buttons. Finally, button (9) is used to find the location of the user if the application did not succeed in finding it automatically. The navigation panel in the bottom of the screen (10) provides navigation details to the user, such as the name of the destination, the estimated time to arrival and the remaining walking distance.

The application does not support a zooming or panning function like traditional mobile map applications: By bringing the device and the map closer, the user can "zoom" into an area of interest. Additionally by moving the device around the map, the user can effectively "pan" into areas of interest. Autofocusing is used to keep the map in clear focus on the mobile screen camera view. Autofocusing also assists in maintaining the recognition of the map throughout use – if the recognition is lost at any point, the icons disappear and the user simply has to re-focus on the map to re-establish recognition.





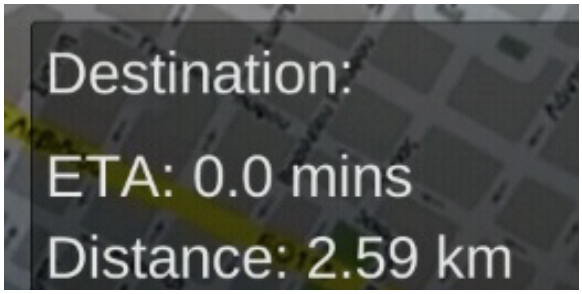


Fig. 2. The HoloPlane Mobile Client Interface

4. Evaluation of POI Visualization Techniques

Our first aim was to explore appropriate visualization techniques in order to help users distinguish and select between the POIs presented by HoloPlane. Our work here is based on a variety of visualization options that is aimed at addressing the visualization of POI categories and POI popularity. For the venue category, since Vuforia allows for the integration of 3D and 2D objects that can be used as markers, we examined both these styles for presenting venue category. We thus created 2D icons (Fig. 3a) and corresponding 3D models representing venue categories. We had two types of 3D models, one being 3D cubes whose sides were textured with the same icons we used for the 2D icons (Fig. 3b), and 3D models of objects representing venue categories (Fig 3c, d). We selected the two 3D icon styles in order to examine how visual complexity affects cognition. The cube representation offers a uniform appearance to the icons (just as in the 2D maps) as well as a visually less complex design, compared to the 3D object models.

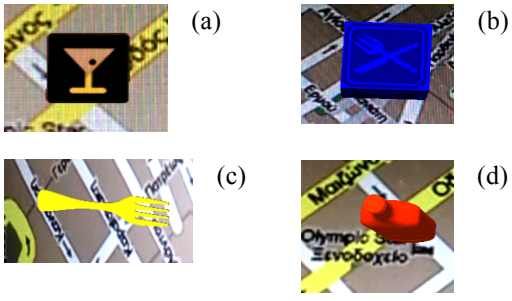


Fig. 3. Examples of HoloPlane marker styles

With regard to venue popularity, we explored two different techniques. One was to alter the size of 2D or 3D markers, thus representing popular venues with larger markers while unpopular venues were represented with smaller ones. The scaling of markers was dynamic and based on their popularity relative to

that of the most popular venue. We implemented a lower threshold for scaling objects, in order to avoid any venues being represented with overly small markers that would make "tapping" the markers on the screen impossible for users. To prevent scaling from creating overly large icons that would obscure proximal icons, we also set an upper threshold of scaling. We also explored a second technique, which was based on the colouring of markers. For this, we explored the use of either discrete colours based on calculated popularity percentiles, hence we had five colours in a spectrum representing the 20th percentile (cyan) to the 80th percentile and above (red). As an alternative, we used a continuous colour spectrum (cyan=most unpopular, red=most popular), allowing for all possible colours in this spectrum based on the popularity of a venue, relative to that of the most popular one (Fig. 4).

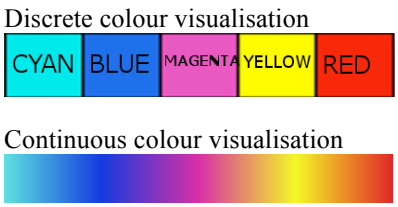


Fig. 4. Venue popularity colour categorisation

Finally, we wanted to explore the efficacy of these techniques in both crowded screens (i.e. with many POIs to visualise) and less crowded ones. These cases represent situations where there are many POI recommendations for the user's query and context, as well as situations where tight filtering (or sparse data) return few results to the user. These are treated separately in our analysis below.

4.1. Preliminary evaluation

To evaluate our prototype, we adopted a scenario based approach comparing different sets of visualizations of the information. For the experiment we provided the participants with a device and a paper map, augmented with fictitious venues to prevent any judgment bias during selection. We generated four sets for three different combinatorial visualizations of points of interest (POIs), using actual POIs returned by the FourSquare API for a specific day and time in our region of interest. The sets were: a) Many 3D POIs, b) Few 3D POIs, c) Many 2D POIs and d) Few 2D POIs. The POI's visualizations cases were: 1) POI size variance (same size and size depending on the popularity of the POI), 2) POI color variance (discrete and gradient depending on the popularity of the POI) and 3) POI shape and model variance (2D shape and 3D cube or 3D object). This setup has as a result thirty-two different visualization combinations.

As a first step, we performed a preliminary experiment via an online questionnaire with 73 participants and presented them with static screenshots of all the visualization combinations in a random order, showing one screenshot for many (18) POIs and one for few (POIs for each visualisation). The visualisations were explained prior to showing them. For each screenshot, participants were asked to respond to the following six questions:

Q1. How many popular venues can you see in this screen?
Q2. How easy is it to distinguish the popular venues in this screen?
Q3. How easy is it to distinguish between the venue categories in this screen?
Q4. How well does the colour of POIs correspond to their popularity?
Q5. How well does the size of the POI correspond to their popularity?
Q6. How would you rate this visualization overall?

Responses to Q2-6 were measured on a Likert scale (1-5) while Q1 was a single choice from the following: 15-20 venues, 21-25 venues, 26-30 venues, 31+ venues. We proceeded with evaluating in a lab experiment, the combinations that scored the highest in each question, which we explored with both few and many POIs. Thus the visualizations to be evaluated in this research are the following (Fig. 5).

V1: *POIs as 3D model having same size and gradient color*
V2: *POIs as 2D shape having same size and gradient color*
V3: *POIs as 3D model having same size and discrete color*
V4: *POIs as 2D shape having same size and discrete color*
V5: *POIs as 3D cube having same size and discrete color*
V6: *POIs as 2D shape having different size and discrete color*

The visualizations in Fig. 5 are shown for the "many POIs" case only. The different visualization styles were programmed into our mobile client, which was able to change its visualization of POIs without

intervention from the researchers, as will be explained next.

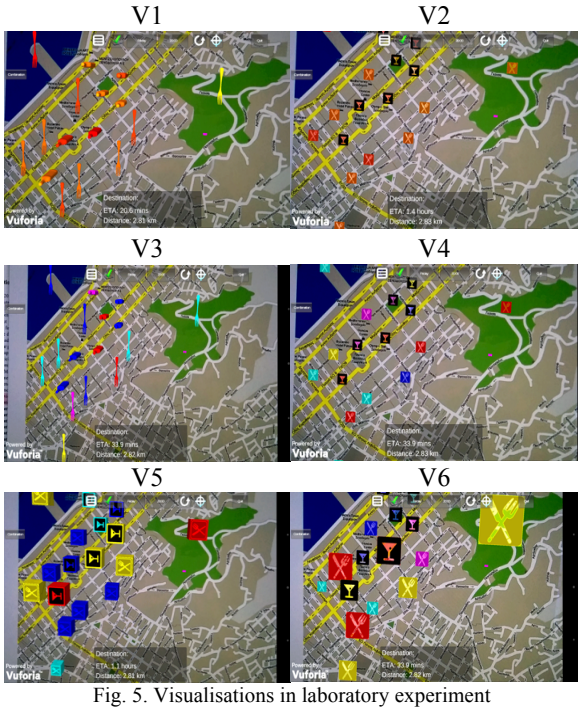


Fig. 5. Visualisations in laboratory experiment

We recruited 20 participants (9 females – 11 male) who were students in our computer science department. Five of them were postgraduate students while the rest being undergraduates. Ten indicated their age category to fall between 18-24 years old, nine between 25-31 years old and one between 32-38 years old. Before starting the experiment, we asked our participants to rate themselves about how familiar they are with using mobile applications and with augmented reality applications on a scale of 1 (very unfamiliar) to 5 (very familiar). Most participants self-reported as familiar or very familiar with mobile apps (22), but unfamiliar or very unfamiliar with augmented reality apps (19).

In order to test the visualizations, we first explained these to participants and then asked the participants to find and select the most popular POI, repeating the same process three times for each visualization (henceforth, this process of submitting three choices for each visualization is called a *session* and each iteration using the same visualization is termed a *sub-session*). Participants were free to view details on as many POIs as they liked before submitting their final choice for each sub-session. Each time the participants made a final choice, the application shuffled the popularity of the POIs and displayed again the same visualization, with POIs remaining in the same geographical location, albeit with different popularities. This simulated the querying of the same dataset under different context (e.g., a restaurant is very popular at 8pm but not popular at all at 8am). After submitting three final selections for the same visualization, the application moved automatically to the next visualization. Each time a participant started the experiment, the application shuffled order with which visualizations were presented, in order prevent any learning effects from affecting the experiment.

4.1.1. Data Collection

We collected all the actions the participants performed while using the prototype during the experiment. We automatically logged the number of POIs viewed by each participant for each sub-session before submitting a final choice, the popularity of the viewed POIs as well as the total time taken to submit a final choice. We also recorded the popularity of the most popular POI for the sub-session, so that it could be compared with the user's actual final choice. At the end of each session, the participants completed a NASA-TLX questionnaire so that we could obtain their subjective workload impression.

4.1.2. Results

This section reports our observations based on the quantitative and qualitative results. The tests reported in this section were chosen according to the outcomes of normality tests on all our variables. We also noted that two of our participants performed the experiment without due attention and seemed to randomly select POIs, exhibiting very low completion times compared to the rest of the users. We therfore excluded the data from these participants from our analysis.

Accuracy of most popular POI detection.

Our analysis begins by examining the participants' ability to find and select the most popular POIs using each visualization. For this, we calculated the relative popularity distance between their final choice and the system's automatically assigned most popular POI in each visualisation. For cases where few POIs are displayed (Fig. 6), participants performed best, exhibiting the smallest average mean distance to the actual most popular POIs popularity with V5F (M=10.22%, SD=11.42%) (lower scores are better). Comparing this visualization to the next best one (V6F), we didn't find a statistically significant difference (Z=-1.503, p=0.133, Wilcoxon). For all

other comparisons, the differences were statistically significant at the $p<0.05$ or $p<0.01$ level (Table 1).

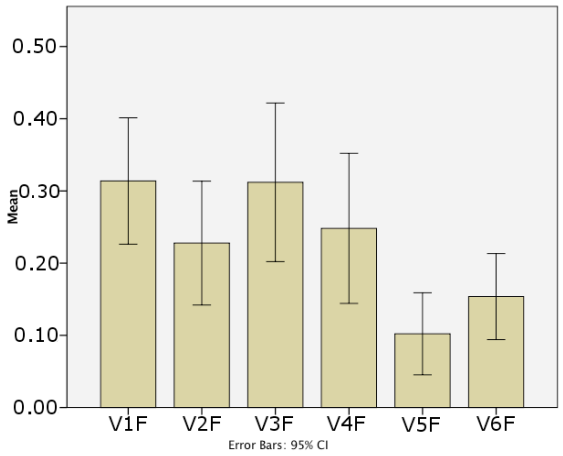


Fig. 6. Mean relative popularity difference of choices compared to most popular POI (few POIs).

Table 1. Statistical analysis results for V5F

	V5F-V4F	**V5F-V3F**	**V5F-V2F**	**V5F-V1F**
Z-value	-3.386	-3.594	-2.133	-3.550
p-value	0.001	0.000	0.035	0.000

For those cases where many POIs are presented (Fig. 7), overall it seems that V6M, V5M and V3M are the three with the best score (M_{V6}=0.14 SD_{V6}=0.11, M_{V5}=0.17 SD_{V5}=0.16, M_{V3}=0.18 SD_{V3}=0.15). Using Wilcoxon signed rank paired tests, we found out that there were no statistically significant differences between the three best performing visualisations (Table 2). At the same time, each of the three best performing visualisations displayed a statistically significant difference to the remaining others.

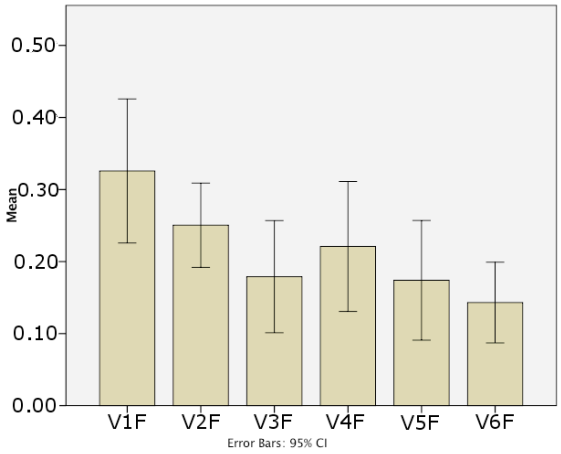


Fig. 7. Mean relative popularity difference of choices compared to most popular POI (many POIs).

Table 2. Statistical analysis results for visualisations with many points (Z, *p* values)

	V2M	**V3M**	**V4M**	**V5M**	**V6M**
V1M	-1.895, 0.058	-3.289, **0.001**	-1.982, **0.47**	-2.940, **0.003**	-2.984, **0.003**
V2M		-2.026, **0.043**	-1.285, 0.199	-2.069, **0.039**	-2.940, **0.003**
V3M			-1.067, 0.286	-0.196, 0.845	-0.457, 0.647
V4M				-1.285, 0.199	-1.329, 0.184
V5M					-0.283, 0.777

Speed of most popular POI detection.

Accuracy has to be paired with speed in order to produce an efficient interface for users. Since selection speed might be affected by the total number of POIs viewed in each visualization, first, we examined participant uncertainty in their choices, by examining the number of POIs viewed in each sub-session before submitting a final choice. For both few and many POI cases, we noted that it was seldom that users viewed more than one POI before making a final choice ($V1_F$=12.96%, $V2_F$=12.96%, $V3_F$=16.67%, $V4_F$=3.70%, $V5_F$=9.26%, $V6_F$=16.67%, $V1_M$=25.93%, $V2_M$=16.67%, $V3_M$=22.22%, $V4_M$=12.96%, $V5_M$=22.22%, $V6_M$=20.37%). Friedman tests showed no statistically significant differences between the visualizations in the number of POIs viewed in the few and many POI cases (Few POIs: $\chi^2(5)=7.880$, $p=0.163$, Many POIs: $\chi^2(5)=3.772$,

p=0.582). Despite observable differences in the same visualisation when comparing their use in few and many POI cases, the presence of more POIs does not seem to affect the number of viewed POIs before the final selection in a statistically significant manner, except for V1 (Z=-2.162, p<0.05, Wilcoxon). From these results we can conclude that our participants generally did not view many POIs before making their selections, although we cannot be sure whether this was because they were mostly certain of their choices while going observing the visualisation outputs, or whether they didn't want to spend too much time exploring options.

To elucidate further, we examined the average time taken in each session with each visualisation (Fig. 8).

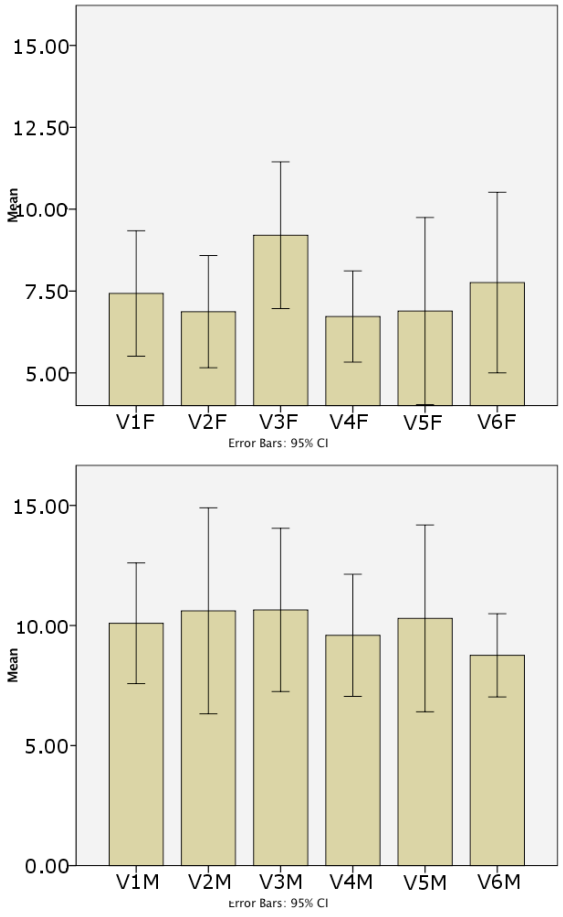


Fig. 8. Average time (seconds) taken to select a POI for few POIs (top) and many POIs (bottom)

Our results show that participants did in fact consider choices carefully by examining the output of visualisations for a considerable time that took several seconds in both few points cases ($V1_F$=7.4s, SD=3.8s, $V2_F$=6.9s, SD=3.4s, $V3_F$=9.2s, SD=4.5s, $V4_F$=6.7s, SD=2.8s, $V5_F$=6.9s, SD=5.7s, $V6_F$=7.8s, SD=5.5s), and many points ($V1_M$=10.1s, SD=5.1s, $V2_M$=10.6s, SD=8.6s, $V3_M$=10.6s, SD=6.8s, $V4_M$=9.6s, SD=5.1s, $V5_M$=10.3s, SD=7.8s, $V6_M$=8.8s, SD=3.5s). Pairwise comparisons between the few and many POI cases for each visualisation showed that the larger number of POIs had a statistically significant effect in the time taken to reach a decision only when using V3 (Z=-2.548, p<0.05) and V5 (Z=-2.026, p<0.05). However, in both cases, Friedman tests showed that participants did not show any statistically significant differences in the time spent for submitting a final choice with any visualisation, in both few and many POI cases (Few POIs: $\chi^2(5)=9.711$, p=0.084, Many POIs: $\chi^2(5)=0.759$, p=0.980). From all quantitative analysis thus, it can be concluded that V5 and V6 are the best choice for helping participants accurately detect the most popular POIs.

Subjective evaluation of visualisations

As discussed, NASA-TLX questionnaires were administered to users after each visualisation session. The participants' responses are shown in Tables 3 and 4 (scale values min=1, max=21). In the following analysis, we did not consider the Physical demand scale as it is not relevant to our work.

Table 3. NASA-TLX scale averages (top cell values) and standard deviations (bottom cell values, italic) for few POIs

	Visualisation					
TLX Scale	V1F	V2F	V3F	V4F	V5F	V6F
M	3.65 *3.00*	3.84 *4.07*	3.65 *2.98*	4.30 *3.89*	3.10 *2.40*	3.25 *3.46*
T	3.75 *3.45*	3.65 *3.50*	3.35 *3.39*	3.45 *2.93*	3.10 *2.38*	2.55 *2.24*
P	4.50 *2.38*	5.10 *4.33*	5.30 *3.89*	4.55 *3.82*	3.70 *3.33*	3.95 *3.66*
E	3.30 *3.53*	4.15 *4.44*	2.60 *2.11*	4.00 *4.00*	2.55 *1.73*	3.50 *2.78*
F	2.80 *2.80*	3.25 *3.60*	2.65 *2.46*	2.90 *3.04*	2.35 *1.93*	2.55 *2.28*

Table 4. NASA-TLX scale averages (top cell values) and standard deviations (bottom cell values, italic) for many POIs

	Visualisation					
TLX Scale	V1M	V2M	V3M	V4M	V5M	V6M
M	4.40 *5.11*	4.20 *4.15*	3.70 *3.63*	3.75 *3.88*	3.00 *2.32*	3.50 *3.71*
T	3.15 *2.92*	4.15 *4.04*	3.15 *2.48*	4.45 *4.41*	3.15 *2.76*	3.40 *3.49*

P	4.45 *4.05*	5.10 *3.61*	4.25 *3.77*	4.65 *3.54*	5.05 *4.19*	4.70 *4.45*
E	3.70 *4.16*	4.00 *4.36*	4.45 *5.15*	3.65 *4.04*	2.95 *1.93*	3.15 *3.03*
F	3.20 *3.72*	3.45 *4.48*	3.80 *3.89*	3.40 *3.93*	2.70 *2.27*	2.80 *3.37*

Despite the differences uncovered in the analysis of quantitative data, Friedman tests failed to reveal any statistically significant differences in any of the visualisations, for both the few POI and many POI cases (Table 5). As a general observation, we note that the scores for the Mental demand, Temporal demand, Effort and Frustrations are low, indicating that participants did not find the tasks overly burdensome with any of the visualisations. Additionally, the Performance scale values are also low (in this case, lower values are best as the scale scoring is inverted), an observation that correlates to the fact that participants did not view many POIs before submitting their choices. This indicates participant confidence in the choices they were making.

Table 5. Friedman test results for NASA-TLX

	Few POIs		**Many POIs**	
TLX scale	$\chi^2_{(5)}$	*p*	$\chi^2_{(5)}$	*p*
Mental demand	5.533	0.354	3.877	0.567
Temporal demand	4.547	0.474	7.044	0.217
Performance	2.443	0.775	2.003	0.849
Effort	8.190	0.146	2.255	0.813
Frustration	2.129	0.831	19.03	0.862

4.1.3. Discussion of results

We noticed here that our participants typically did not explore many POIs before making a selection, using all visualisations. This shows that participants had confidence in their choices, particularly given the fact that they spent non-trivial amounts of time to examine their available choices. Given the absence of statistically significant differences between visualisations in considering the time taken to reach a decision, we turn to the accuracy with which participants managed to identify the most popular point with each visualization. Here, V5 is the clear winner in situations where few POIs are displayed on the map. It is also one of the top 3 performing visualization techniques when many POIs are displayed, without exhibiting a statistically significant difference to the other two. Despite these differences, participants did not appear to notice significant differences in the visualisations from a subjective perspective. Hence, our recommendation for designers are:

- Context information relating to recommendation (in our case, popularity), is best encoded into markers as discrete colours from a palette of choices (V5, V6).
- If using 3D models as POI markers, then these are best when displayed as cubes of the same size (V5). No further support for context representation is required.
- If using 2D icons as POI markers, then further support for context representation is required by additionally manipulating the marker size (V6).

5. Group navigation using augmented paper maps

In this section, we discuss our field experiment using real tourists, who we subjected to a group navigation task in small teams. Our purpose here was to explore the dynamics of group navigation with our augmented paper map prototype and measure the extent to which differences emerge compared to the use of a traditional mobile navigation application (Google Maps). The experiment described here is focused solely on navigation aspects, hence we did not ask our users to find and select POIs to visit. These were pre-selected for them by the researchers and the origin and destination POIs were the only visual elements, together with the route, shown on the mobile screen.

5.1. Experiment design

Our participants were 23 undergraduate engineering students from various disciplines (14 male, 9 female), from 17 European countries, who were visiting the city of Patras for a summer school. Their ages ranged between 18 and 26 years old and none had previous experience with mobile AR applications. All participants mentioned familiarity with navigation applications, with 40% stating frequent use and 17% indicated always using just a mobile application while visiting a new place. We

found a low preference for fixed city maps (e.g., wall-mounted) and paper maps (22% in both cases) compared to mobile navigation apps. To establish thus a baseline that would be representative of our participants' usual behaviour, we chose to compare our prototype to the most preferable navigation aid for our participants, i.e. a mobile navigation app and not a paper map. Hence we selected the familiar navigation tool installed on all Android devices, i.e. Google Maps (GM). For the field experiment we provided participants with four devices of equivalent capabilities in terms of processor speed and screen size (LG Nexus 4 and Nexus 5 and Samsung S3 and S4), which all ran our application with good performance. In order to test our prototype in navigation tasks, we established two routes of equal complexity in terms of turns and walking distance (Fig. 2), requiring approximately 10 minutes of walking time from a person familiar with the area. We let participants split themselves into 8 groups, allowing friends to work together to better simulate real tourist groups – the first four groups completed the first route using the HoloPlane AR prototype and proceeded to GM navigation for the second route. This order was reversed for the remaining four groups. Each team was accompanied by a researcher who knew the routes and was able to provide help if the team did not succeed to find the destination. Finally, in each team, one user volunteered to control the device and map (where used), while the other two participants were termed as "companions" and were instructed to ask for control of the device and map, if they so desired. This setup is representative of situations where one person assumes the navigator's role, typically because they own the device. As stated previously, HoloPlane is designed to be used with any simple printed map. For our experiment, we provided participants with a colour printed map from the Google Maps website that shows the experiment area at a scale of roughly 1: 18055 (zoom level 16). This is the smallest scale at which Google shows names for all streets and not just major ones. Furthermore, this scale allows the map to depict as wide an area as possible, maintaining label readability for the users. We selected an A4 print size, to represent a typical situation for users who might have printed a map at home before travelling, or during their stay (e.g., at Internet café), as few users would typically have access to a large format printer such as A3 or larger.

5.2. Data Collection

We collected GPS positioning data for each team. The researchers, who accompanied each team, also noted the number of times participants stopped to consult the application and make a route choice during navigation. At the end of each navigation task we asked each participant to complete a NASA TLX questionnaire, so that we could obtain their subjective workload impression. We also asked them to complete two validated questionnaires for each system: a Brief Mood Introspection Scale (BMIS) [14] in order to measure mood and a User Experience Questionnaire (UEQ) [13] for their overall experience.

5.3. Results

This section reports our observations based on the quantitative and qualitative results. The tests reported in this section were chosen according to the outcomes of normality tests on all our variables.

Quantitative measures.

In Fig. 9, we show the participants' walking behaviour during navigation, which is visualized through a heatmap-based depiction of GPS traces. We report this data as recorded by the device GPS without statistical significance analysis, since the number of teams was too small to provide an adequate sample size for statistical significance.

Overall teams took less time to navigate with GM (M_{gm}=896.96s, SD_{gm}=295.93s, M_{hp}=1093.93s, SD_{hp}=319.10s). However with GM they made more stops to consult the tool (M_{gm}=8.75, SD_{gm}=9.77, M_{hp}7.88, SD_{hp}=3.41). We measured the length of pauses they made during navigation, i.e. periods longer than 5 seconds where the speed was less than 1km/h. There were fewer such periods with GM (M_{gm}=10.86, SD_{gm}=3.31, M_{hp}=17.86, SD_{hp}=7.22) which lasted also less time (M_{gm}=273.88, SD_{gm}=241.19, M_{hp}=316.87, SD_{hp}=178.0). Finally, in terms of distance covered, this was less with GM (M_{gm}=690.80m, SD_{gm}=81.01m, M_{hp}=842.79m, SD_{hp}=192.8m).



R1-GM



R1-HP



R2-GM



R2-HP

Fig. 9. Participant Routes and heatmapped GPS traces. The red segments show where participant speed was less than 1 km/h

Participant Workload Assessment

At the end of each navigation task, we issued each participant with a NASA-TLX questionnaire to obtain their subjective ratings of their experience with each navigation tool. The overall results are summarized in Fig. 10. Overall it can be seen that GM was rated better than our prototype (a lower score is better), with the exception of physical effort. The latter is expected, as the routes were carefully chosen to present equal levels of walking difficulty and length. Concerning the remaining five variables, a statistical significance in the difference of means was only found for effort to complete the task, using a paired-sample T-test (M_{gm}=7.61, SD_{gm}=4.878, M_{hp}=10, SD_{hp}=3.357, $p<0.05$) and performance, using a Wilcoxon signed rank test (M_{gm}=3.43, SD_{gm}=3.287, M_{hp}=6.26, SD_{hp}=4.826, $p<0.05$). Overall thus it appears that the GM tool led to the expenditure of less effort to complete the navigation task and participants felt more successful using it.

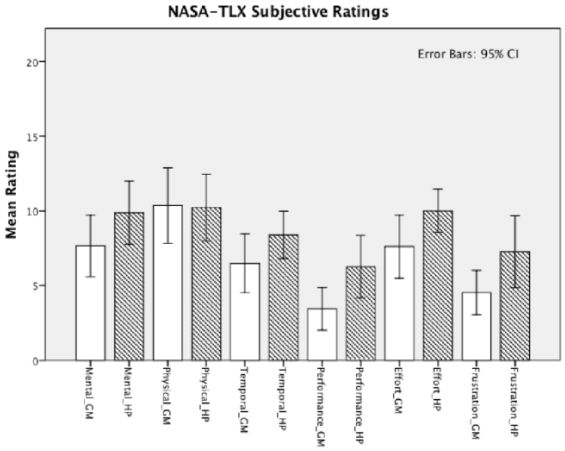


Fig. 10. Subjective Workload Assessment

Participant Affective State

Using the BMIS questionnaire at the end of each task, we asked participants to give us insight to their affective state during the tasks. This questionnaire contains 16 adjectives describing affective state. Before letting the participants answer the questionnaire, we explained in detail each adjective, in order to be sure that they fully understood the choices and their meaning. The analysis of the user responses was made on the Calm - Arousal and Unpleasant - Pleasant axes, and is depicted below in Fig. 11. It can be generally seen that the participants' experience was rated positively in terms of pleasantness and that participants felt averagely aroused during the navigation tasks.

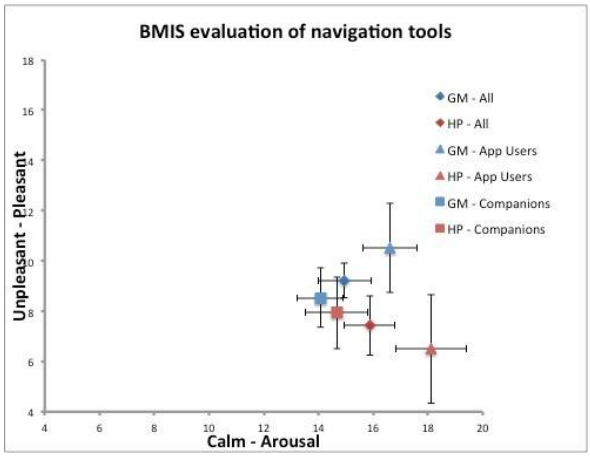


Fig. 11. Affective state during navigation tool use

Further analysis reveals that when considering all users, no statistically significant differences using Wilcoxon signed rank tests for the two navigation tools, on either the Unpleasant-Pleasant (M_{gm}=9.22, SD_{gm}=4.69, M_{hp}=7.44, SD_{hp}=5.67) or the Calm-Arousal axis (M_{gm}=14.96, SD_{gm}=3.28, M_{hp}=15.87, SD_{hp}=4.38). We went further by breaking up the users according to their roles (app users and companions) and analyzing the respective data. We did not find any statistically significant differences using Wilcoxon signed rank tests in either axis for any of these user categories. Given the previously found statistically significant difference in performance, we conclude that while the participants believe they fared worse with the HoloPlane prototype, nevertheless, their experience was just as pleasant as with GM.

Participant User Experience.

At the end of each navigation task, we asked each participant to complete the User Experience Questionnaire, in order to obtain a measure of their assessment of each navigation tool (Fig. 12). The questionnaire generally assumes a positive appraisal on each dimension if the mean exceeds 0.8, or a negative appraisal if the mean is less than 0.8. Analysis with Wilcoxon signed rank tests reveals that statistically significant differences appear only in the dimensions of perceived *Efficiency* (M_{gm}=1.978, SD_{gm}=0.170, M_{hp}=0.578, SD_{hp}=0.875, $p<0.01$), *Dependability* (M_{gm}=1.674, SD_{gm}=0.82, M_{hp}=0.924, SD_{hp}=0.89, $p<0.05$), *Stimulation* (M_{gm}=0.728, SD_{gm}=1.047, M_{hp}=1.467, SD_{hp}=0.728, $p<0.01$) and *Novelty* (M_{gm}=-0.63, SD_{gm}=1.297, M_{hp}=1.609, SD_{hp}=0.856, $p<0.01$). These outcomes for the *Stimulation, Efficiency* and *Dependability* are in line with the outcomes from our previous questionnaires. The observed difference in *Stimulation* measures is somewhat unexpectedly in disagreement with the parity observed in the Arousal-Calm axis earlier. However, a more careful inspection of the wording of the UEQ adjectives used to measure on the positive scale for this axis, uncovers that these imply a level of *engagement*, instead of measuring affective state (*valuable, exciting, interesting, motivating*). Finally, there is clear indication here that our participants considered HoloPlane to present significant novelty.

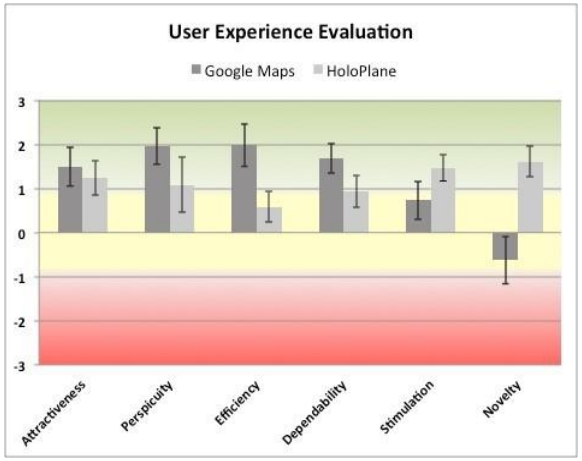


Fig. 12. User experience during navigation tool use

Other observations

When observing participant bodily configuration, we noticed a more relaxed approach with the AR tool, compared to "squeezing in" to view the device instructions when using GM, an observation also made in [16]. In Fig. 13, we show several examples of use of the HoloPlane prototype. In these, the shared use of the hybrid working space is evident in several collaboration examples: In the first (Fig. 13a), the "navigator" has control of both the paper map and the device. Companions are gathered around the map, paying attention to the printed surface which is clearly visible and intelligible to all, while the screen of the device is used only by the navigator. His role here is to communicate what he sees on the device, to the companions, so that a shared understanding can be achieved. Communication is verbal, since both the navigator's hands are occupied. In the second example (Fig. 13b), the "navigator" controls the device, while one of the companions is holding the map. Here, the "navigator" is seen to be pointing on the map, in order to communicate to the companions his knowledge in a more comprehensible manner. This mode of communication is more direct and helps companions

understand more easily what the navigator sees. Finally, in Fig. 13c, we note that the communication of spatial awareness is initiated by the companion, who is holding the paper map and at the same time pointing to a location on it. At the same time, the navigator is trying to understand the companion's communication and match it to what is represented on the device screen. This example shows that the hybrid system allows for more active participation in the navigation task by all group members.

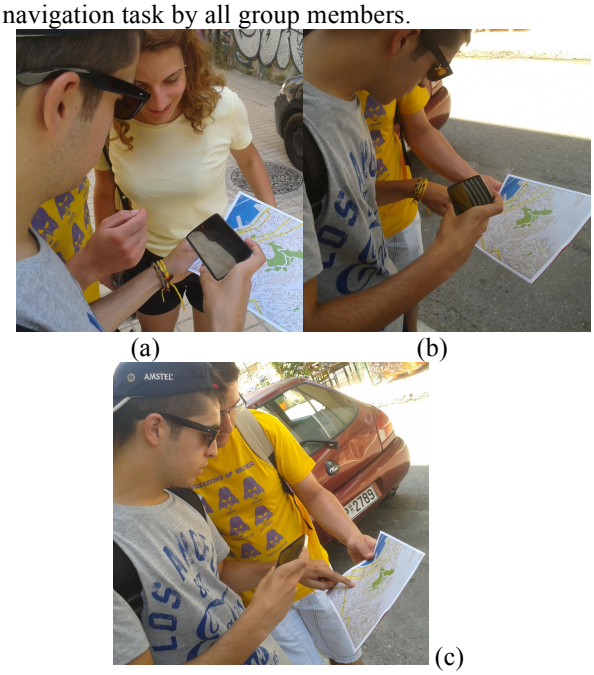


Fig. 13. Group behaviour during use of the Holoplane hybrid interface

The next figure (Fig. 14) shows some instances of the navigation task, during use of the Google Maps interface. Here it is easy to observe that the planning task is made much more difficult for all users, since the screen real-estate is quite small and participants have to gather tightly to see what is displayed. Not all participants are able to point to the screen in order to communicate their understanding, hence limiting their ability to make a contribution to the planning (Fig. 14a). During transit to established waypoints, the companions often resigned to being simple followers (Fig. 14b). Here, the companion on the right is talking to the navigator, since they were able to plan the route together previously, leaving the female companion unable to contribute to the planning. The female companion, adopts a passive mode since she did not participate in the planning stage, and is seen to be walking just ahead of the group, keeping an ear out for the navigator's next instruction. This is evident also in Fig. 14c, where the female companion is simply looking around. The navigator is ahead of the group on his own, trying to determine the group's whereabouts, while the male companion is trying to visually match the surrounding location to the printout of the navigation target given to the group.

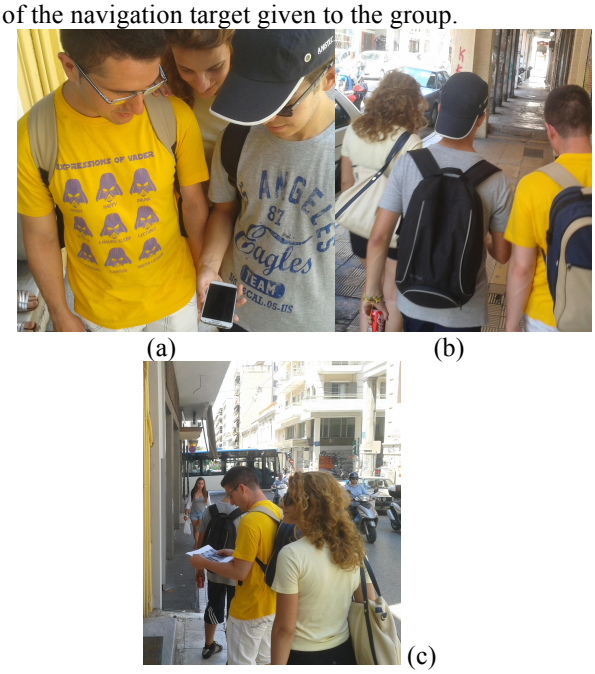


Fig. 14. Group behaviour during use of the Google Maps interface

6. Discussion and Future Work

This paper introduces a novel system that allows users to generate their own printable map targets and to associate these with contextual information on POIs, retrieved by querying popular social networks using augmented reality. Our evaluation into this system presents two contributions: First, we evaluated the design of context-related POI markers for augmenting paper maps, with the purpose of facilitating the process of finding recommended venues. Secondly, we investigated the use of augmented paper maps, with a specific focus on the dynamics of group navigation.

Regarding the design of context-related markers, we examined a range of visualisation options and found that while users did not subjectively perceive differences between these options, recommendations for design choices can be made based on the participants' actual behaviour during the use of these visualisations. Our experiment was based on a lab study, which of course carries limitations that are inherent to this type of exploratory activity. We had initially postulated that variable marker sizes would

prove beneficial to users as they would facilitate easier "tapping" of markers in order to explore the options available to users, however our lab experiment showed that users did not really explore the POIs shown to them in order to make a choice, and hence it was the marker colours and design which played a role in assisting choices. Marker size seemed to play a role only when using 2D icons. It can be argued that because this was a lab experiment with fictitious POIs, actual user behaviour in terms of exploring choices might be different. A field trial comparing the effectiveness visualisation techniques would be a natural next step to complete the evaluation process. However, such a process would be difficult to organise, due to the volume of visualisation options and a requirement for capturing a large number of real tourists. Deployment of our service as a real-world application might be able to yield better insights.

Regarding navigation, our evaluation was based on the use of validated questionnaires whose use is not widespread in the field of mobile HCI. This approach contrasts previous research in [15] and [16] whose findings are based on the analysis of qualitative interviews. Yet, our preliminary evaluation did not find any significant performance advantages of augmenting a paper map for navigation, a result that is completely in line with [15] and [16]. This outcome provides indication that the questionnaire-based approach has merit and can be used effectively in the place of qualitative interviews, where the danger of researcher bias in the analysis of results is significant. Another similarity with [16] is that when observing participant bodily configuration, we noticed a more relaxed approach with the AR tool, compared to "squeezing in" to view the device instructions when using GM (Fig. 13).

The reason why no advantages were observed with the AR interface may relate to the size of the augmented paper map. We selected a relatively small printed area (A4) to represent a typical situation of users printing their own maps. Perhaps a larger shared map might make the magic lens interface more usable, as suggested by [9], although there, maps were fixed on to a wall surface, where as in our scenario users have to be able to conveniently hold the map. Hence, while providing a larger printed map might make its augmentation more usable, it might detract from its key benefit (i.e. portability and manipulability). A further consideration for performance is item density: In our situation, the item density was very low and included just two POIs and the route. As per [21], it can be expected that our users might have focused more on the paper map than the magic lens, hence preventing the system from achieving its performance potential. Further tests with different item densities (e.g., routes with multiple waypoints) would be needed to verify any effects.

As indicated by the Stimulation axis in the UEQ, our participants felt more engaged as group members with the HP system than GM, where a single user takes on the role of the navigator and collaboration is hindered, as the small screen limits the information space. The reported level of engagement might be an effect of the high perceived novelty of the system, since both axes (Stimulation & Novelty) relate to hedonic quality perception. However, the UEQ Novelty axis has been found not to correlate with the Stimulation axis in other research [22]. As a side effect of increased engagement with the navigation task, the acquisition of spatial knowledge for all users might be improved for users as per [25], but further tests would be needed.

It is encouraging that participants found the AR tool just as attractive as the standard navigation tools. The issues of mental workload and efficiency appraisals can be attributed to the novelty and unfamiliarity of our application to users.

To this end, we are hoping to conduct further, more extensive trials to eliminate familiarity factors from the results. Furthermore, given that augmented maps can be used as a collaboration tool, our future research will also encompass the use of our AR tool with public displays of maps.

7. References

[1] Adams, A. and Cox, A. L., Questionnaires, indepth interviews and focus groups. *In Research Methods for Human Computer Interaction*, Cairns (2008), P. and Cox, A. L. (eds.). Cambridge University Press, Cambridge, UK, 17–34.

[2] Bouwer, A., Nack, F. and Evers, V. Towards support for collaborative navigation in complex indoor environments. *In Proc. CSCW* (2011), 601-604.

[3] Brown, B., and Chalmers, M, Tourism and mobile technology. *In Proc. ECSCW* (2003), 335-354

[4] Chittaro, L. Visualizing information on mobile devices. *Computer*, 39(3), 2006, 40-45.

[5] Čopič Pucihar, K., Coulton, P. and Alexander, J. The use of surrounding visual con- text in handheld AR: device vs. user perspective rendering. *In Proc. CHI* (2014), 197-206.

[6] Delikostidis, I., van Elzakker, C. P., and Kraak, M. J. Overcoming challenges in developing more usable pedestrian navigation systems. *Cartography and Geographic Information Science* (2015), 1-19.

[7] Dünser, A., Billinghurst, M., Wen, J., Lehtinen, V. and Nurminen, A. Exploring the use of handheld AR for outdoor navigation. *Computers & Graphics* 2012, 36, 8: 1084-1095.

[8] Elias, B., and Paelke, V. User-centered design of landmark visualizations. *In Map-based mobile services*, 2008, (pp. 33-56). Springer Berlin Heidelberg.

[9] Grubert, J., Pahud, M., Grasset, R., Schmalstieg, D. and Seichter, H. The utility of Magic Lens interfaces on handheld devices for touristic map navigation. *Pervasive and Mobile Computing* (2014), 18, 88103.

[10] Hile, H., Vedantham, R., Cuellar, G., Liu, A., Gelfand, N., Grzeszczuk, R., and Borriello, G. Landmark-based pedestrian navigation from collections of geotagged photos. *In Proc. of the 7th international conference on mobile and ubiquitous multimedia* (2008) (pp. 145-152). ACM.

[11] Huang, H., and Gartner, G. A Technical Survey on Decluttering of Icons in Online Map-Based Mashups. *In Online Maps with APIs and WebServices*, 2012, (pp. 157-175). Springer Berlin Heidelberg.

[12] [*Left blank for review*]

[13] Laugwitz, B., Held, T., and Schrepp, M. Construction and evaluation of a user experi- ence questionnaire. *In HCI and Usability for Education and Work, Holzinger, A. (ed.)*, 2008, , (pp. 63-76). Springer, Heidelberg, Germany.

[14] Mayer, J. D. and Gaschke, Y. N. *The experience and meta-experience of mood*, in: Journal of Personality and Social Psychology, 55, 1988, pp. 102-111.

[15] Morrison, A., Mulloni, A., Lemmelä, S., Oulasvirta, A., Jacucci, G., Peltonen, P. and Re- genbrecht, H. Collaborative use of mobile augmented reality with paper maps. *Computers & Graphics* (2011), 35, 4: 789-799.

[16] Morrison, A., Oulasvirta, A., Peltonen, P., Lemmela, S., Jacucci, G., Reitmayr, G. and Juustila, A. Like bees around the hive: a comparative study of a mobile augmented reality map. *In Proc. CHI* (2009), 1889-1898.

[17] Müller, J., Jentsch, M., Kray, C. and Krüger, A. Exploring factors that influence the combined use of mobile devices and public displays for pedestrian navigation. *In Proc. NordiCHI* (2008), 308-317.

[18] Norrie, M. and Signer, B. Overlaying paper maps with digital information services for tourists. *In Proc. ICTT* (2005), 23-33.

[19] Raita, E. User interviews revisited: identifying user positions and system interpretations. *In Proc. NordiCHI* (2012), 675-682.

[20] Reilly, D., Mackay, B., Watters, C. and Inkpen, K. Planners, navigators, and pragmatists: collaborative wayfinding using a single mobile phone. *Personal and Ubiquitous Computing* (2009), 13, 4: 321-329.

[21] Rohs, M., Schöning, J., Raubal, M., Essl, G. and Krüger, A. Map navigation with mobile devices: virtual versus physical movement with and without visual context. *In Proc. ICMI* (2007), 146-153.

[22] Schrepp, M., Hinderks, A., and Thomaschewski, J. Applying the User Experience Questionnaire (UEQ) in Different Evaluation Scenarios. *In HCI* **8** (2014) (pp. 383-392).

[23] Stroila, M., Mays, J., Gale, B. and Bach, J. Augmented transit maps. *In Proc. WACV* (2011), 485-490.

[24] Teevan, J., Karlson, A., Amini, S., Brush, A. J., and Krumm, J. Understanding the importance of location, time, and people in mobile local search behavior. *In Proc. of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services* (2011) (pp. 77-80). ACM.

[25] Willis, K. S., Hölscher, C., Wilbertz, G., and Li, C. A comparison of spatial knowledge acquisition with maps and mobile maps. *Computers, Environment and Urban Systems* **33**(2) (2009), 100-110.