A Dialogue Based Mobile Virtual Assistant for Tourists: The SpaceBook Project

Phil Bartie[1], William Mackaness[2], Oliver Lemon[3], Tiphaine Dalmas[2], Srini Janarthanam[3], Robin Hill[2], Anna Dickinson[2], Xingkun Liu[3]

1 University of Stirling; 2 University of Edinburgh; 3 Heriot-Watt University

Accepted refereed manuscript of:

Bartie P, Mackaness W, Lemon O, Dalmas T, Janarthanam S, Hill R, Dickinson A & Liu X (2018) A dialogue based mobile virtual assistant for tourists: The SpaceBook Project, Computers, Environment and Urban Systems, 67, pp. 110-123.

DOI: 10.1016/j.compenvurbsys.2017.09.010

A Dialogue Based Mobile Virtual Assistant for Tourists: The SpaceBook Project

Abstract

Ubiquitous mobile computing offers innovative approaches in the delivery of information that can facilitate free roaming of the city, informing and guiding the tourist as the city unfolds before them. However making frequent visual reference to mobile devices can be distracting, the user having to interact via a small screen thus disrupting the explorative experience. This research reports on an EU funded project, SpaceBook, that explored the utility of a hands-free, eyes-free virtual tour guide, that could answer questions through a spoken dialogue user interface and notify the user of interesting features in view whilst guiding the tourist to various destinations. Visibility modelling was carried out in real-time based on a LiDAR sourced digital surface model, fused with a variety of map and crowd sourced datasets (e.g. Ordnance Survey, OpenStreetMap, Flickr, Foursquare) to establish the most interesting landmarks visible from the user's location at any given moment. A number of variations of the SpaceBook system were trialled in Edinburgh (Scotland). The research highlighted the pleasure derived from this novel form of interaction and revealed the complexity of prioritising route guidance instruction alongside identification, description and embellishment of landmark information – there being a delicate balance between the level of information 'pushed' to the user, and the user's requests for further information. Among a number of challenges, were issues regarding the fidelity of spatial data and positioning information required for pedestrian based systems – the pedestrian having much greater freedom of movement than vehicles.

Keywords
Location based service; Spoken Dialogue System; Viewshed; Virtual City Guide

1 Introduction

Technology has long supported tourists in experiencing the city from trip planning, to finding public transport information, to providing navigation assistance, to post-trip reminiscing (online photo sharing and blogs). The smartphone has revolutionised how travellers personalise their travel experiences but an ongoing concern of the smartphone platform is that finding relevant information on screen distracts the user from their environment - a conspicuous and somewhat hazardous activity. What is required is a more concealed technology that supports intuitive interaction but does not come between the tourist and their enjoyment of the city. This paper reports on the SpaceBook Project (EU ref: 270019), which focussed on designing a wearable technology that delivers relevant information (information push), and responds to user questions (information pull) while the tourist explores the city on foot. In order that it leave the tourist both hands-free and eyes-free, the system was entirely dialogue based using only speech input and output. The system was able to provide navigation guidance as well as identify landmarks in view (e.g. statues, buildings, parks) by modelling visibility in real-time based on a Digital Surface Model (DSM) built from LiDAR data. While there are many approaches to informing a mobile user, (Google glass, haptic interfaces), to our knowledge SpaceBook was the first system to rely solely on natural language speech and text-to-speech responses to support in situ navigation and exploration of the urban environment. While there have been industry led speech-based virtual agents (e.g. Apple's Siri, Microsoft's Cortana), SpaceBook is unique in modelling context by calculating the pedestrian's field of view in order to provide situated speech-based dialogue to support both navigation and exploration in the truest sense of the word.

The system was evaluated by 42 people on the streets of central Edinburgh (a busy area crowded with visitors and traffic and with a wide variety of geographic features and topography). The evaluation had four main aims: (1) to establish the performance of continuous Automatic Speech Recognition (ASR) in a noisy outdoor environment; (2) to model object visibility in real-time and in conjunction with social media data (e.g. Flickr, Foursquare) in order to determine useful landmarks to assist in navigation tasks or information push; (3) to evaluate pedestrian level positioning and tracking in the urban environment; (4) to determine the optimal balance in the delivery of 'pushed' information and user requests alongside navigation instructions.

2 Background

The increase in processing power of mobile devices has enabled a new generation of mobile spatial interaction (MSI) (Carswell, Gardiner, & Yin, 2010) allowing users to interact more easily with relevant digital information in their surrounding environment. For example IT solutions exists that enable users to navigate and find out the name of a landmark using an

Augmented Reality application (Chung, Pagnini, & Langer, 2016; Gu & Gu, 2016; Liarokapis, Mountain, Papakonstantinou, Brujic-Okretic, & Raper, 2006; Narzt, et al., 2006). Smartphones have become the most suitable candidate for MSI because of their combination of 1) small form, 2) positioning capabilities and other sensors (acceleration, barometer, gyroscope), 3) data transfer via mobile networks, and 4) sufficient battery power for a day of use.

When people explore an unfamiliar environment, particularly a cityscape, they spend a lot of time looking around. Most tourist systems require the user to interact with a graphical interface, which distracts them from appreciating the city or to paying attention to obstacles in their path (Heuten, Henze, Boll, & Pielot, 2008). Speech, on the other hand is one of the most natural forms of interaction, and is particularly suitable when a user is carrying out tasks that occupy their view (e.g. driving, walking) or makes physical demands (e.g. opening doors, carrying, physically aiding others). The ambition of SpaceBook was to focus on this speech interaction in order to build a hands-free, eyes-free application that enabled users to explore and be guided around a city. The system used speech as its only user interface, without any use of the phone's display, such that the phone remained concealed in a pocket or bag. Interaction was via headphones and microphone. Speech interfaces in industry such as Siri, Cortana, Alexa, and Google Assistant, normally respond to only a single utterance from the user, such as a web search or command, with minimal follow-up conversation. In contrast, the SpaceBook system deployed a Spoken Dialogue System where multi-turn sequences of interaction are employed in a long-running conversation with the user, lasting many minutes. Such dialogue based interaction can better reflect the collaborative nature of exploratory learning, with well understood interaction benefits (Cai, Wang, MacEachren, & Fuhrmann, 2005). However this conversational style of interaction (Lemon, 2012) is challenging because it has to perform tasks and meet the user's goals across long sequences of turns, maintaining an accurate representation of the context at all times. Standard research on dialogue systems focusses on single tasks, such as restaurant search (Young, Gasic, Thomson, & Williams, 2013) or flight booking, where the user's location is static. The dialogue system for SpaceBook had to manage a much more challenging situation, with multiple tasks (e.g. navigation, points-of-interest, question-answering) within a dynamic location-based system, and so it constitutes one of the most complex spoken dialogue interfaces yet created. The main novel contribution of the SpaceBook dialogue system is its location-based Interaction Manager (see section 4.7) (Srinivasan Janarthanam, et al., 2013) which handled multiple conversational threads (Lemon & Gruenstein, 2004). The system also used a continuous speech recogniser, which was always listening to the user, rather than a push-to-talk system or one triggered by `hot-words' such as "OK Google" or "Alexa", which are used in current commercial systems.

Location Based Services (LBS) enable more effective system interactions by automatically including the user's location in the search. The pioneer of LBS was Cyberguide (Long, Aust, Abowd, & Atkeson, 1996) which could calculate its location indoors using infrared beacons and outdoors using the Global Positioning System (GPS), providing location customised information to tourists. The system demonstrated that mobile computing was able to usefully adapt information delivery based on location and place histories, offering an alternative to a human tour guide. A similar system for blind pedestrians was designed by R. G. Golledge, Klatzky, Loomis, Speigle, and Tietz (1998), and evaluated by Loomis, Golledge, and Klatzky (1998), that proposed speech based input coupled with spatialized sound to convey information about the immediate environment. A wide variety of location aware applications followed, including GUIDE a virtual guidebook (Davies, Cheverst, Mitchell, & Friday, 1999), GEONOTES for attributing space using virtual tags (Espinoza, et al., 2001), way-finding applications (A. J. May, Ross, & Bayer, 2005; Andrew J. May, Ross, Bayer, & Tarkiainen, 2003), friend finding (Strassman & Collier, 2004), urban gaming (Benford, et al., 2006), and EASYGO for public transport information (Gartner, et al., 2007).

The initial uptake of LBS by the population was fairly slow, which can be partially attributed to poor user experiences, service unreliability, and a lack of perceived ownership benefits (Chincholle, Goldstein, Nyberg, & Eriksson, 2002). Furthermore many potential users were concerned about issues of privacy (Duckham & Kulik, 2006) and security (Cahill, et al., 2003). Their established ubiquity has arrived with mobile computing platforms that are continuously geolocated using solutions such as Global Navigation Satellite Systems (GNSS) and WiFi positioning, with freely available location aware applications (e.g. Google Maps, TripAdvisor, AirBnB). Most applications use a simple measure of distance to determine geographical relevance when filtering information. For example, the closest park determined in Euclidean space, or the nearest supermarket using network space. People, however, often refer to items in vista space (Montello, 1993), defined as the region visible from a location (Figure 1).

Figure 1: Map of Vista Space - the regions visible from a specified location (green dot) are highlighted in yellow

Augmented Reality applications (e.g. Layar, Wikitude) overlay name labels on an image captured from a smartphone's camera to notify the user of surrounding features, which is a development of the point-to-query interaction pioneered by Geowand (Egenhofer, 1999), where the user's position and the orientation are used to control the data filter. A more recent example of this which includes a visibility filter, is Zapp (Meek, Priestnall, Sharples, & Goulding, 2013) which runs on a smartphone allowing the user to discover information about things in their line of sight, such as the geology of a distant hill. Prior to this visibility modelling was included in the Edinburgh Augmented Reality System (EARS) (Bartie & Mackaness, 2006) which through a speech interface supported pedestrian urban exploration. EARS announced landmarks as they came into view, enabling the appropriate audio keyword so that the user could ask for more details about any previously announced landmark by name via the speech interface. The system supported free exploration but not navigation, and the visibility information was pre-calculated for 86 selected landmarks within the pilot study region (Edinburgh city). There are other examples of systems which avoid visually distracting the user, such as through haptic feedback (Heuten, et al., 2008) or abstract sound (S Brewster, 1997; S. Brewster, 1998) but these tend to be limited in what they can communicate and are not intuitive as they require the user to learn how the interface presents information.

3 Design Aspects

Various forms of information are required to support the process of exploration which is a complex dynamic process. Urban exploration is about having the capacity to roam freely, retaining a sense of where you are, whilst acquiring spatial knowledge and a sense of place (Reginald G. Golledge, 1992). Wayfinding is just one component of exploration; much has been written on the role that landmarks play as confirmatory cues, or at key decision points along a route (Duckham, Winter, & Robinson, 2010; Richter & Winter, 2014; Sorrows & Hirtle, 1999). Augmented information supporting this process often comes in map form (with its associated cognitive effort). However in dialogue based systems the cognitive effort comes in providing useful descriptions of objects (buildings, statues, street names) that are readily seen, and unambiguous in the field of view. Redundant (superfluous descriptions) and verbose descriptions reduce ambiguity but leave less time for other interactions whereas a careful choice of landmarks (i.e. choosing highly salient landmarks for which there are no distractors) facilitates brevity. For example if there is only one castle, then 'the castle' is sufficient – notwithstanding the need for a shared prototypical understanding of what a castle in a city might look like.

Field based Wizard-of-Oz experiments (Kelley, 1983) were conducted in order to understand the mediating role of a dialogue based virtual assistant and the level of detail required to support exploration and wayfinding. Wizard-of-Oz experiments are where the user is exposed to a system that gives the illusion of being a working system when in reality it is operated by a human (wizard) hidden 'behind the screen' (Alce, Wallergard, & Hermodsson, 2015). In addition to information relevant to wayfinding, the experiments also examined how information snippets describing landmarks could be 'pushed' to the user, and how subsequent user requests for more detailed information could be responded to ('pull' information).

Most challenging of all was the observation that conversations with participants were 'interleaved' with push/pull information together with time critical wayfinding information. For example:

System (push): 'On your right you can see the castle'
User (pull): 'Ah yes, can you tell me more? '
System (push-wayfinding): 'turn right at the junction'
User (pull): 'which junction? '
System (push):'the castle is open to the public'

System (push-wayfinding): ‘the junction next to the Bank of Scotland’

The difficulty is knowing when to push, in what order to respond to requests, whilst retaining the flexibility to adjust responses according to their relevance (whether the object under discussion is still in view or not). This prioritising was handled through an Interaction Manager that was able to prioritise multiple requests and responses (Section 4.7). It became apparent that prioritisation needed to be modelled based on the immediacy of the task, together with both physical and social qualities (Section 4.5). These observations were reflected in the subsequent design of the system and its implementation.

4 System Components

SpaceBook was designed using a client-server architecture with a number of micro-services at different sites communicating via the internet. The following section gives a brief overview of each component and their integration (Figure 2).

Figure 2: SpaceBook system components and connections

4.1 Phone Application (Client Side)

The phone application carried out two tasks concurrently, 1) the relaying of audio over a voice channel (user’s voice to the automatic speech recognition engine, and synthetic speech back from the text-to-speech (TTS) engine, and 2) the transmission of position and accelerometer data to the Pedestrian Tracker (PT) via a 3G data channel. It is worth noting that not all mobile network providers offer the ability to connect voice and a full rate data channel simultaneously, and some experimentation was required to find a suitable network provider. Attempts to send both compressed audio and location data across the 3G data connection (using transmission control protocol (TCP) or user datagram protocol (UDP) were problematic as during longer street experiments, especially in busy areas, sections of audio would be delayed or garbled (with UDP) causing automatic speech recognition (ASR) errors. Therefore the audio channel was used as it provided a much more robust connection for sending the user’s audio to the ASR. A street style headset with an attached noise cancelling microphone was worn by the user, connected via a splitter cable to the phone’s audio jack, enabling close-miking to improve ASR success rate.

The client application was designed to run on Android phones as a background service so that positional data would be streamed even when the screen was turned off. The positional data (i.e. latitude, longitude, speed, orientation) were sent across the 3G network on each GNSS update event (1Hz), along with a summary of the step rate calculated from the on-board 3-axis accelerometer. These were received by the PT and processed in order to improve locational accuracy through map matching algorithms (Section 4.6).

4.2 Automatic Speech Recognition (ASR)

The speech recognition was handled by Nuance 9.0 (www.nuance.com) and FreeSWITCH (an open source telephony framework – freeswitch.org). using a grammar-based language model (Saksamudre, Shrishrimal, & Deshmukh, 2015). The audio channel was kept open with the ASR running continuously (i.e. without a "push to talk" button as used in Google Voice Search, S-Voice, and other mobile applications). The grammar and vocabulary included entity names (e.g. streets, shops, statues) and entity types (e.g. park, café, hotel, supermarket) within the test area (Edinburgh City), as well as the names of prominent people that may be the subject of Question-Answering inputs (e.g. Mary Queen of Scots, Harry Potter). The grammar consisted of approximately 80 rules, covering user navigation goal inputs, Question-Answering inputs, visibility statements, and general dialogue-management inputs. Such structuring provides a framework in which text strings can be broken down into their constituent forms. The grammar fragment presented in Table 1 is a standard hand-crafted rule set that constrains the speech recogniser to only recognise vocabulary and structures as defined in the grammar (McTear, 2002). This technique helps to increase accuracy of speech recognition in noisy environments.

Table 1:Example Grammar Model

- (i [(want) (need)] to [(get) (go) (walk) (get directions to)] to PLACE)
- (what is [TYPE THING] ?(famous for))
- (?okay i ?can see [([a the] LANDMARK) TYPE_OR_STREET] ?now)
- (?(can you) say that again ?please)
- (how far is it ?(to PLACE))

4.3 Semantic Analysis (SA)

The SA module translated natural language utterances into a machine-interpretable meaning representation language (MRL). The SA was trained from hand-coded utterances captured during the Wizard-of-Oz experiments. From the experiments 17 dialogues were collected, resulting in 1906 annotated utterances used to train the SA. These gave a sense of user expectations and the types of demands they might typically make of a digital tourist guide. Figure 3 gives examples of the MRL output generated from the SA component, and full details of the SA can be found in Vlachos and Clark (2014).

Figure 3: Sample dialogue annotated with Meaning Representations

4.4 City Model (CM)

The City Model (CM) acted as the central repository for information about the city, containing spatial representations of features as well as attributes and functions to process requests (e.g. shortest path), in PostgreSQL with PostGIS. It consisted of a wide variety of sources including Ordnance Survey's (OS) Master Map which provided geographical land use details (e.g. building, pavement), OS PointX and OpenStreetMap point data for occupancy details including postal address, feature use and name (e.g. name=Nile Valley, type=Restaurant). To allow flexibility in how the information was stored, the database was vertically partitioned, such that the *isA* table held details on feature types at various hierarchical levels (e.g. book shop => household, office, leisure => retail), while the *isNamed* table stored the various occupant names (i.e. Blackwell's Bookshop). Using this schema a single physical entity (e.g. building) could be linked to many uses and names, to model the relationship between physical structure and the multiple occupying businesses. This provided the flexibility to easily add additional names and types in order to accommodate a wider range of user requests. A side effect was that it complicated the process of deciding the most appropriate name to refer to a building during an information push event. Therefore, the most popular name was determined by measuring web saliency, given as the number of matches on crowd sourced media including Twitter, Flickr, and Foursquare. The feature type was also modelled such that salient use types (e.g. bank, post office, food franchises) received high rankings.

It was also necessary to customise the name selection based on the user's approach angle so that the system's reference to a building matched the user's view. This was done by linking given occupant street address information to the road network, such that each occupant point was assigned a direction. In the example shown in Figure 4, a user approaching the marked *building* from [a] would be notified of '*Caffé Nero'*, while the same building would be referred to as '*Blackwell's*' when the user approached from [b] matching the shop frontages. Such an arrangement provided the flexibility to refer to the building using either name, in response to a user information 'pull' event.

Figure 4: Example of Building with multiple entrances and shared utility (Café and Bookshop). Label Selection depends on Approach Direction
(MasterMap data, Ordnance Survey © Crown copyright. All rights reserved OS))

The CM also provided shortest path calculations based on a pedestrian accessible network constructed from OpenStreetMap (OSM) data. The pgRouting extension (pgRouting.org) was used to solve the shortest path which was then turned into a set of natural language navigation instructions using stored procedures. Four added descriptions were used to embellish the description of each segment of the route: the hill gradient for each segment was based on a terrain model (enabling comments such as: 'head up the hill'), the path type from OSM was stored in the CM (e.g. street, bridge, steps). Network segment sinuosity was derived using angle measurement thresholds, and junction type (node degree) appropriate to the approach angle (e.g. T, Y, X). An estimation of which roads were more well-known was also introduced based on the number of Flickr images taken on each street, and the number of Foursquare check-ins. This gave a proxy to rank the popularity of visits to each street section. Such information could be used to ask the user if they knew how to navigate to that well-known road from which the system could take over navigation, thereby reducing the number of navigation instructions needed. Figure 5 is an example of the generated output table from the CM, with the corresponding route map. Such information supported delivery of natural phrasing (e.g. 'take the right fork and go slightly down hill along the street')

Figure 5: Route instructions including topography, junction shape, and a metric for better known streets
(MasterMap data, Ordnance Survey © Crown copyright. All rights reserved OS; © OpenStreetMap contributors)

The CM also included other features to aid the Interaction Manager (Section 4.7) such as fuzzy text string matching (trigram and phonetic matching), and the ability to generate initial directional guidance. This was used to clarify which direction to start walking when the GNSS direction value was not trustworthy as GNSS can only calculate orientation from movement history. The magnetometer could not be utilised because the phone was not held flat, but was instead in their pocket or bag, resulting in a very noisy output. The solution to orientating the user was to refer to well-known nearby landmarks and ask the user to keep those on their right or left as they set off (e.g. "keep Blackwell's Bookshop on your right"). This ensured they did not have to back track once the GNSS was able to derive a good direction heading from the trajectory history.

4.5 Visibility Engine (VE)

The Visibility Engine (VE) modelled the user's vista space (Montello, 1993) by accessing a DSM built from LiDAR data, giving a 2.5D representation of the city including buildings, vegetation, and land surface elevation (Figure 6). Before the DSM could be used for visibility modelling it required 'cleaning' to remove the shapes of cars and buses from the roads captured by the LiDAR. This was done by passing a focal minimum kernel across raster cells within the road region as defined by the OS Master Map polygons. Particular care was taken on bridges to ensure elevations from the lower road were not incorrectly transferred to the road above.

Figure 6: Digital Elevation Models (a surface model and a terrain model)

The visibility model was implemented using a sweep algorithm, which scanned a 360-degree region of 5000m radius around the user. To ensure responsiveness the sweep algorithm was parallelised to use all available cores on the server in order to return sub-second results. It was considered worthwhile to calculate the visibility in all directions for two reasons: (1) so that SpaceBook could draw attention to any interesting features to the side, in front or behind the user; (2) so that calculated results could be cached and used again no matter what the revisit approach direction. Cached viewshed results could be retrieved within 20ms using a quadtree index.

The results from the cell visibility were summarised per feature based on OS Master Map polygons. The zonal statistics included the distance to the closest and furthest visible part of each feature, the ID for foreground objects blocking the view, field of view occupied, and statistics on the vertical extent and façade area showing for the feature. Bartie, Reitsma, Kingham, and Mills (2010) provide further details on how these fields were calculated. The Digital Terrain Model (DTM) was also used to measure the vertical extents of buildings and façade areas. The visible extent of an object was calculated by comparing the interception height of the lines of sight against the DTM. In the example in Figure 7, we see the full vertical extent (A1 to A2) of building A is in view, while only a small portion (h2) at the top of building B is visible.

Figure 7: Visible Extent for a feature based on difference between DSM and DTM values

As well as modelling individual polygons the visibility model could return site summaries, such that a group of polygons could be considered as a single entity. For example, Edinburgh Castle consists of museums, open spaces, armouries, and cafes, but should be addressable as a single entity for both push and pull events.

Figure 8: Edinburgh Castle - considered as a single site or as 326 separate polygons (© OS Master Map)

When combined with the information from the City Model the Visibility Engine could automatically identify key landmarks along a generated route, which formed useful anchor points for turn instructions, as well as confirmation along the route to reassure the user they were heading in the right direction. A combination of factors were used to calculate the saliency of a building: its proximity to a decision point (i.e. a turn along the route being followed), the amount visible (façade area), the number of times seen along the route, and its presence in social media (Flickr, Foursquare). In this manner, it was possible to rank buildings along a proposed route. This proved critical to the efficient selection of salient features that would be most readily identified and recognised when giving route following instructions (Figure 9).

Figure 9: Important Landmarks ranked along a route (blue line)
(MasterMap data, Ordnance Survey © Crown copyright. All rights reserved OS))

4.6 Pedestrian Tracker (PT)

The Pedestrian Tracker (PT) module combined the GNSS position and sensor data (e.g. accelerometer) with spatial data from the City Model to calculate the most likely position of the pedestrian, thus improving upon the raw GNSS output. The user's trajectory needed to be analysed in near real-time without the opportunity to look ahead and implement techniques such as corner matching (Meng Yu, 2006). The PT generated two outputs: a probability array for the user's location, and a snapped road centreline position. The array was used to determine the most likely location of the user from which to calculate the user's viewshed, while the snapped centreline position gave a more robust solution for navigation purposes based on the road and track network dataset. Analysis of the GNSS speed and sensor values enabled a degree of confidence to be given to each location solution. For example if a location update resulted in a jump of 15 metres, but the accelerometer data indicated the pedestrian had only taken three steps then the reported location would be assigned a low confidence level.

Typically in outdoor environments GNSS (e.g. GPS, GLONASS, Galileo) is suitable for positioning, however in urban canyons (i.e. between tall buildings) the direct line of sight to the satellites may be occluded, and multipath signals reach the receiver after being bounced off nearby surfaces resulting in positioning errors (Mountain & Raper, 2001; Raper, Gartner, Karimi, & Rizos, 2007). SpaceBook was designed for pedestrians and consequently the user was typically on the pavement next to tall buildings, encountering greater satellite occlusion than vehicles on the road. Pedestrians are less restricted in their movements than vehicles, are able to turn on the spot (GNSS is unable to track turning while not moving forward) and are not limited to following networks (instead crossing roads, plazas and parks). All of this makes tracking challenging.

To improve the situation a smartphone that could receive both GPS and GLONASS was used (a Samsung Galaxy Note), thereby increasing the number of potential satellites in view at any time and improving positioning robustness (Fantino, Mulassano, Dovis, & Lo Presti, 2008). Figure 10 shows a comparison trial conducted along a narrow Edinburgh street, which demonstrates the advantages of using a phone able to harness GPS and GLONASS systems compared to only GPS.

These include:

- faster initial (cold) position lock (typically under 10 seconds, compared to around 45 seconds) as indicated by the stuttered purple dots (GPS only) near the start location.
- keeping a position lock for longer in urban canyons, and even indoors on some occasions. a better location solution across a range of environments. This can be observed in the shape of the trajectory near the junction turns in Figure 10. The GPS+GLONASS track more closely follows the right angle turns, while the GPS phone has a heavily filtered (rounded) pathway output. Similar experiences were noted by Mattos (2011), who reported that position solutions were 2.5 times better when using GLONASS in addition to GPS.

Figure 10: GPS (purple line) vs GPS+GLONASS (green) trajectories, together with typical street view.

(MasterMap data, Ordnance Survey © Crown copyright. All rights reserved OS)

Given SpaceBook's ambition to model vista space it was important that the reported user location did not fall on roof tops, as a horizontal positional error of only a few metres could result in very different viewshed results calculated from roof top positions. To overcome this a Pedestrian Accessibility Model (PAM) was developed which represented the likelihood that a pedestrian could occupy a space based on the land use type. Pavements were considered to have a higher occupancy probability than roads, and open spaces slightly less than pavements but more than road regions. The PAM is shown in perspective view and map view in Figure 11, where the 'elevation' value represents the probability of user occupancy (taller = less likely to be occupied). Therefore all buildings appear as equal heighted 'loaves', and pavements appear as 'gutters'. From this perspective, the user's location is analogous to a marble rolling around this 'loaf world' such that the user's location is gently pushed towards the most likely region. This is useful if the reported location is on the roof of a building, since the aspect of the slope indicates the most likely direction of movement required for the correction.

Figure 11: Pedestrian Accessibility Model (PAM)
(a) perspective and (b) map views of for the surfaces most likely to be occupied by the pedestrian

As the user moved around a region a candidate space probability map was generated by the PT which included the most likely position of the user in real-time. Figure 12 shows an example of this where the GNSS reported position (green dot) is on a road, and the PT has corrected the location to the pavement (red dot) taking into consideration the PAM, user's speed, reported facing direction, and previous location.

Figure 12: Location Probabilities in the Candidate Space around the Reported GNSS Location
(MasterMap data, Ordnance Survey © Crown copyright. All rights reserved OS))

4.7 Interaction Manager (IM)

The Interaction Manager (IM) was the central module receiving the user's location coordinates from the PT and the semantics of the user's utterances via the SA module. It was responsible for interleaving three tasks:

1. Navigating the user: The IM handled navigational requests in three stages: (1) identifying/suggesting destinations, (2) presenting the route instructions, and (3) 'closing' the task. First, the user's request was analysed, and if it was deemed a navigational request the IM would query the CM for a route to the destination. Route instructions were presented in situ to the user at every decision point, using visible landmarks as references whenever possible, until the user reached the destination. The IM detected and revised the route plan if the user walked in the wrong direction. On approaching the destination, the IM presented the relative direction to reach the goal.

2. Pushing information about Points of Interest: The IM queried the CM in order to create a list of additional items (such as points of interests - PoI) that were close to the user. The IM then queried the Question Answering (QA) module for information on the PoI and presented the response segments to the user. This allowed the user to ask for more information on an item that interested them (Section 4.8).

3. Exploration using Question Answering (QA): The system was able to detect open questions (Questions that could not be posed to the CM). Open questions from the user such as "Who is David Hume?", "Tell me more about the castle", "What is that?" were sent to the QA module, which returned a text string that the IM presented to the user via the TTS engine (Janarthanam, et al., 2012).

Speech based interfaces are restricted to serial communication (user says something, SpaceBook says something back), therefore the IM had to prioritise the information delivery based on what was most important at any given time. SpaceBook therefore had to balance concurrent tasks of wayfinding guidance, user questions, and landmark push information. These were allocated different priorities so that the time critical information was pushed to the user as soon as possible. For example if a user is approaching a key decision point along a route, a wayfinding instruction needs to be prioritised over an information push about a historical building (Janarthanam & Lemon, 2014). Thus the dialogue actions in the queues were pushed to the user according to the following order of priority:

Priority 1. Dialogue control (repeat request, clarifications)
Priority 2. Responding to user requests
Priority 3. System initiated navigation task actions
Priority 4. Responses to User-initiated QA actions
Priority 5. Point of Interest (PoI) Push actions

Dialogue actions in the queues were revised periodically to reflect changes in context. Obsolete dialogue actions were removed to avoid pushing entities that were no longer relevant, and to allow other important dialogue actions to be pushed at appropriate times.

4.8 Question Answering (QA)

The Question Answering (QA) service found information relevant to open requests from the user such as descriptions of things, biographical information and additional information about anything they have heard (pull behaviour). Their requests were also in response to IM queries (push behaviour) about the user's surroundings. QA extracted answers from a custom index built from Wikipedia and the Scottish Gazetteer (Gittings, 2017). It handled dialogue history and contextual references (anaphoric and deictic expressions).

The QA service had its own type of analysis based on open domain techniques for textual search (Li & Roth, 2002; Mikhailian, Dalmas, & Pinchuk, 2009). When a question was deemed out of scope by the IM (i.e. not answerable by the city model), the question was passed to QA – in essence becoming the content provider – drawing upon unstructured data sources in order to provide additional information. The core functionality provided 'pull' information (responses to a question posed by the user) and 'push' information. These were answers to questions proactively generated by the IM. For example the IM might identify from the city model a point of interest that was nearby or in the field of view which then required description.

This functionality was delivered via four modules: 1) question classification, 2) focus extraction, 3) co-reference checking and resolution, and 4) search. The first module classified the question into one of four types: (i) out of scope (i.e. questions not covered by QA such as ‘Where is the Royal Mile?’ or ‘How much is an entrance ticket?’), (ii) biography (i.e. questions about people such as ‘Who is John Knox.. what is he famous for?’), (iii) description (‘What is haggis?’, ‘Tell me about John Knox House’),and (iv) next segment (i.e. where the user seeks further information on a previous topic, eg ‘Tell me more..’).

The second module (focus extraction) pinpointed the focus of the question (eg ‘John Knox’, ‘Statue’). Requests for information fall broadly into two types, anaphoric where the question relates to prior discussion (e.g. where prior discussion was about David Hume and the user requests: ‘tell me more about him?’), or deictic – in which the question is related to some object in the field of view (e.g. the user spots the national museum and asks ‘What is that?’). So the third module was required to resolve whether the question was anaphoric or deictic. Anaphoric questions require QA to keep a track of previous transactions in order to resolve ambiguities and prevent repetition.

The fourth module conducted the search. The answer from the search came from one of two sources: the Gazetteer for Scotland (Gittings, 2017) or Wikipedia (en.wikipedia.org). The Gazetteer for Scotland details information on 325 points of interest and 391 descriptions of famous people. Relevant entries from Wikipedia were scoped by only using links connected with Edinburgh (10,898 entries). Experiments with WordNet 3.0 (which provides dictionary style definitions for common names – (Miller, 1995)) did not prove to be sufficiently useful because of its generic nature, it created more ambiguities than it resolved, and was thus excluded. The search initially checked for anaphoric candidates and if none were found then the focus became deictic. Deictic questions could be answered because the VE was able calculate which PoIs dominated the user’s field of view at any given instant and included the ability to filter the results by pre-defined characteristics (for instance by type: such as statue, church, park). The QA component could then select the top PoI from the city model, and the deictic question analysed to see if it satisfied the constraints expressed by the focus (e.g. ‘What is that church?’ or if the pedestrian is in front of a statue, ‘Who is it?’).

4.9 Natural Language Generation (NLG) and Text to Speech (TTS)

The Natural Language Generation component took content planning input from the IM and realized it in English. It used dictionaries for City Model names and type constants encoding grammatical and morphological features in order to construct sentence text which was passed to the speech engine. CereProc (http://www.cereproc.com/) was used as the Text-to-Speech Engine, with a Scottish female voice called ‘Heather’. The only changes made were minor adjustments to the pronunciation of certain place names (e.g. “The Pleasance”). The audio file output from this was sent to the client over the audio phone channel via freeSWITCH.

5 SpaceBook Evaluation in Edinburgh, Scotland

There are considerable challenges in performing usability evaluations on non-traditional interactive systems (Dünser & Billinghurst, 2011). In essence, SpaceBook was built on experiments to understand the intent of the user (WoZ), together with formative and summative evaluations. While there are a range of situated mobile learning platforms that share the same intention as SpaceBook, their focus on a visual interface makes comparison with intentionally concealed technology problematic. Furthermore, measuring the efficacy of SpaceBook via time based exercises ran contrary to the nature and idea of roaming and exploring the city. Instead evaluation focused on the user's response to the novelty of dialogue based interaction (Srini Janarthanam, et al., 2013) and the shared nature of task execution between the human and the device (Carroll, 1991).

5.1 Evaluation

Evaluation consisted of comparing three configurations of SpaceBook in order to assess the effectiveness of its various functions. Three variants were created: (System 1) a Multi-threaded Interaction Manager and Visibility Engine (the information from the visibility engine being used as a basis for both navigation and information pushing); (System 2) same as system 1 but with a simple single-threaded Interaction Manager that did not prioritising dialogue actions; (System 3) was the same as system 1 but without the Visibility Engine, therefore navigation and information pushes used only the proximity of a PoI to identify landmarks and points-of-interest.

A 7 point Likert-type scale (1 Strongly Disagree, 4 neutral, 7 Strongly Agree) was used to evaluate the navigation and discovery tasks, followed by a post experiment debrief. The details of what was asked, and why, are provided in a report online (http://www.spacebook-project.eu/pubs/D6.2.2.pdf).

5.1.1 Participants

42 participants were recruited: 24 students (8 male/16 female; mean age 23, age range:16-40) and 18 people over 50 (10 male/ 8 female; mean age 62.4, age range: 52-76). All participants rated themselves "Fit and able", and could walk for 90 minutes and cope with steep and uneven ground; they were all native speakers of English, with a range of accents (including Northern Irish, New Zealand, and Indian). Participants participated in a two-hour session and were paid £25 irrespective of the outcome of the experiment.

5.1.2 Task protocol

Participants attended for a two-hour session that started when the participant met the researcher who explained the study, administered a demographic questionnaire (age, fitness, familiarity with smart phones) and explained the informed consent form. The participant and the researcher then walked to the start of a route where the researcher fitted the SpaceBook system, and started it, and repeated the experiment instructions, giving the participant the opportunity to ask questions. The participant was then given the first task, and asked to imagine that they were a tourist exploring the city with no constraint on their time. It was explained that SpaceBook would tell them about things that it thought they might be interested in, and they could ask it questions about things that they wanted to know about. As participants carried out the tasks, the researcher followed at a distance in order to observe, whilst avoiding interacting with the subject. The participant was instructed to complete the task on their own, using the system, and to only talk to the researcher if they were completely stuck. At the conclusion of the exercise a questionnaire was completed, and this was combined with researcher observations, and telemetry collected by the system (velocity profile, push/pull instructions, clarification requests, etc).

5.1.3 Routes

The experiment comprised three co-located legs chosen for their diversity of views and route complexity (Figure 13).

Figure 13: The three routes in the city of Edinburgh
(© OpenStreetMap contributors)

5.2 Results

After all participants had completed the routes they were asked to complete a feedback form. The main conclusions from the experiments were as follows:

- 91% of the navigation tasks were completed successfully
- ASR struggled to perform well in noisy environments with false positives and accuracy issues. The IM was able to handle some misrecognitions but 43% of the QA pull requests were misrecognised due to ASR errors.
- The older cohort of users found the specificity of content to be too generic, and would have preferred richer information of greater depth.
- Users did not like the system prioritising its content over their questions (this resulted in information requests going unanswered)
- 32% of the QA interactions were related to visual spatial co-references (deictic) (e.g. what is on my left?), 10% to dialogue co-references (anaphoric) (e.g. when was it built?), 4% were proximal in form (e.g. what landmarks are nearby?), and 54% were without a co-reference. ASR recognition errors will likely have skewed these results.

- Confidence in the system was eroded during periods of silence (e.g. along featureless sections of road). It is argued that such silences should be filled with confirmatory queues to instil confidence that the system was tracking the user's position correctly.
- When presented with a wide city vista more careful consideration should be given to describing features in view, otherwise there was potential for the system to overload the participant with information and descriptions.
- Overall users felt least in control of System I (VE+Multi-threaded), and found System III (no VE, multi-threaded IM) the most helpful. It is argued that these feelings may have arisen from the verbosity of the system and the challenges of trying to prioritise large amounts of 'push' information which resulted in delays to responses to 'pull requests'.
- The general acceptability of a dialogue based system such as SpaceBook was reflected in the 63% of respondents who said they would use it again.

It was interesting to note that users quickly became reliant on the system with an expectation of detailed guidance, for example: '*No information was given on whether to cross the street*', '*I was not told to cross the road at almost any point*'. In some cases the complexity of the environment warranted greater guidance – making comments such as: '*at the complex junction … much more info is needed to guide user where to go, as that junction is difficult to cross with complex traffic flow*'. Users sometimes required more detailed descriptions of landmarks with comments such as: '*I didn't know what the Black Watch monument looked like, and SpaceBook gave no description…*', '*The system should consistently provide landmark descriptions*', and '*Non-description of buildings is difficult for tourists'*.

Confidence in the system was eroded if they could not identify reported landmarks indicated by comments such as: '*At Milnes Court it was using landmarks that couldn't be seen.*', '*When SB is talking about things you can't see, or identify, it has an unsettling effect.*'

5.3 Future Improvements

The users were asked to consider what improvements should be made to the system after completing all of the navigation and exploration tasks. The most requested change was to improve the ASR recognition rate in outdoor noisy environments. Another request was for more detailed information on the points of interest beyond that of Wikipedia and the gazetteer. Users liked the descriptions added to help identify landmarks (e.g. the building with a green dome), but would like more detailed descriptions for more city objects.

Figure 14: Improvements to SpaceBook (out of 42 participants)

In some instances, the sentence snippets from QA were not easy to understand. Wikipedia entries were difficult to read by the TTS because of the long sentences and use of parentheses. Use of Simple English Wikipedia (simple.wikipedia.org) helped solve this problem, but it provided improved readability for only a limited portion of the content. Work on text simplification (Woodsend & Lapata, 2011), in particular for TTS, would likely improve user comprehension ratings. Alternatively, techniques providing generated answers (as opposed to extracted answers) could be explored as a way to improve content delivery.

In terms of potential improvements to the system, it was difficult to find a balance between the sequencing, timing of delivery and the amount of information as it related to different tasks. For example when the users were in the middle of a navigation task, they were confused by the interjection of pushed information. As one user put it: '*you worry that SpaceBook has forgotten what you said you wanted to do and is telling you about other things instead'*. There were also complex configurations when the user was navigating to a required point whilst listening to the answer to a pull request, followed by some pushed facts. It was not always obvious to the user that the system had gone from one topic of interest to another. Thus there would be merit in adding topic switching utterances (e.g. "*On the subject of David Hume*…"). Deictic questions were also difficult to handle given uncertainties relating to user position and viewshed, and the number of possibly relevant PoI in an area. Such a situation yielded several satisfying candidates – consequently SpaceBook had much to talk about! The need was felt for the system to further pin-point the user interest through dialogue in order to better manage the flow of information. Future versions may explore allocation of types of information between male/female synthesised voices.

The experiments revealed how critical the timing of information delivery was, for example when push information based on a viewshed trigger was delayed by even a few seconds it could result in the user being told that an item was in view which was now occluded. This could be overcome by calculating the duration an object was likely to remain in view, and by using the IM to check the validity of items at the time they are announced. Such lags in the receipt of information leads to confusion and a degree of mistrust. Trust is critical in such systems where the balance of decision making is shared between the participant and the machine. More broadly, linking information to human location and activity raises issues of privacy - something common to LBS (Bridwell, 2007). Any commercial development of SpaceBook would need to consider ways of protecting the locational privacy of users.

6 Conclusion

SpaceBook demonstrated a pedestrian based virtual guiding system running on a smartphone client connected over a 3G network to a set of services. The system relied on a speech only interface so that the user could maintain an eyes-free hands-free experience while exploring the city. It was found that generally most people that used the system enjoyed the experience, and would like to use such a similar system in the future. The biggest issues were the quality of continuous ASR in a noisy outdoor environment (misrecognition sometimes making the interaction frustrating), managing large volumes of information (in open vistas with many potential push candidates), and in prioritising and balancing push/pull information.

7 Acknowledgements

The research leading to these results has received funding from the EC's 7th Framework Programme (FP7/2011-2014) under grant agreement no. 270019 (SpaceBook project).

8 REFERENCES

Alce, G., Wallergard, M., & Hermodsson, K. (2015). WozARd: A Wizard of Oz Method for Wearable Augmented Reality Interaction-A Pilot Study. *Advances in Human-Computer Interaction*.

Bartie, P., & Mackaness, W. A. (2006). Development of a speech-based augmented reality system to support exploration of cityscape. *Transactions in GIS, 10*, 63-86.

Bartie, P., Reitsma, F., Kingham, S., & Mills, S. (2010). Advancing visibility modelling algorithms for urban environments. *Computers Environment and Urban Systems, 34*, 518-531.

Benford, S., Crabtree, A., Flintham, M., Drozd, A., Anastasi, R., Paxton, M., Tandavanitj, N., Adams, M., & Row-Farr, J. (2006). Can you see me now? *ACM Transactions on Computer-Human Interaction (TOCHI), 13*, 100-133.

Brewster, S. (1997). Using Non-Speech Sound to Overcome Information Overload. . *Displays - Special Issue On Multimedia Displays, 17*, 179-189.

Brewster, S. (1998). The design of sonically-enhanced widgets. *Interacting with Computers, 11*, 211-235.

Bridwell, S. A. (2007). The dimensions of locational privacy. *Societies and Cities in the Age of Instant Access, 88*, 209-225.

Cahill, V., Gray, E., Seigneur, J. M., Jensen, C. D., Chen, Y., Shand, B., Dimmock, N., Twigg, A., Bacon, J., English, C., Wagealla, W., Terzis, S., Nixon, P., Di Marzo Serugendo, G., Bryce, C., Carbone, M., Krukow, K., & Nielsen, M. (2003). Using trust for secure collaboration in uncertain environments. *IEEE Pervasive Computing, 2*, 52-61.

Cai, G., Wang, H., MacEachren, A. M., & Fuhrmann, S. (2005). Natural conversational interfaces to geospatial databases. *Transactions in GIS, 9*, 199-221.

Carroll, J. M. (1991). *Designing interaction: Psychology at the human-computer interface* (Vol. 4): CUP Archive.

Carswell, J. D., Gardiner, K., & Yin, J. (2010). Mobile visibility querying for LBS. *Transactions in GIS, 14*, 791-809.

Chincholle, D., Goldstein, M., Nyberg, M., & Eriksson, M. (2002). Lost or Found? A usability evaluation of a mobile navigation and location-based service. *Mobile HCI, 2411*, 211-224.

Chung, J., Pagnini, F., & Langer, E. (2016). Mindful navigation for pedestrians: Improving engagement with augmented reality. *Technology in Society, 45*, 29-33.

Davies, N., Cheverst, K., Mitchell, K., & Friday, A. (1999). Caches in the air: disseminating tourist information in the Guide system. *Second IEEE Workshop on Mobile Computing Systems and Applications* pp. 11-19). New Orleans, Louisiana: IEEE.

Duckham, M., & Kulik, L. (2006). Location privacy and location-aware computing. In J. Drummond (Ed.), *Dynamic & mobile GIS: investigating change in space and time* pp. 34–51). Boca Raton, FL: CRC Press.

Duckham, M., Winter, S., & Robinson, M. (2010). Including landmarks in routing instructions. *Journal of Location-Based Services, 4* 28-52.

Dünser, A., & Billinghurst, M. (2011). Evaluating Augmented Reality Systems. In B. Furht (Ed.), *Handbook of Augmented Reality* pp. 289-307). New York, NY: Springer New York.

Egenhofer, M. (1999). Spatial information appliances: A next generation of geographic information systems. *1st Brazilian workshop on geoinformatics, Campinas, Brazil*.

Espinoza, F., Peterson, F., Sandin, P., Nystrom, H., Cacciatore, E., & Bylund, M. (2001). GeoNotes: Social and navigational aspects of location-based information systems. In S. Shafer (Ed.), *Ubicom 2001* pp. 2-17). Atlanta, Georgia: Springer.

Fantino, M., Mulassano, P., Dovis, F., & Lo Presti, L. (2008). Performance of the proposed Galileo CBOC modulation in heavy multipath environment. *Wireless Personal Communications, 44*, 323-339.

Gartner, G., Cartwright, W., Peterson, M. P., Pun-Cheng, L. S. C., Mok, E. C. M., Shea, G. Y. K., & Yan, W. Y. (2007). EASYGO: A public transport query and guiding LBS. *Location Based Services and TeleCartography* pp. 545-556). Springer Berlin Heidelberg.

Gittings, B. (2017). The Gazetteer for Scotland.

Golledge, R. G. (1992). Place recognition and wayfinding: Making sense of space. *Geoforum, 23*, 199-214.

Golledge, R. G., Klatzky, R. L., Loomis, J. M., Speigle, J., & Tietz, J. (1998). A geographical information system for a GPS based personal guidance system. *International Journal of Geographical Information Science, 12*, 727-749.

Gu, R., & Gu, J. L. (2016). Research on the Key Techniques of Augmented Reality Navigation. *Proceedings of the 2016 International Conference on Automatic Control and Information Engineering (Icacie), 64*, 161-163.

Heuten, W., Henze, N., Boll, S., & Pielot, M. (2008). Tactile wayfinder: a non-visual support system for wayfinding. *Proceedings of the 5th Nordic conference on Human-computer interaction: building bridges* pp. 172-181). ACM.

Janarthanam, S., & Lemon, O. (2014). Multi-threaded interaction management for dynamic spatial applications. *EACL 2014* (p. 48). Gothenburg, Sweden.

Janarthanam, S., Lemon, O., Bartie, P., Dalmas, T., Dickinson, A., Liu, X., Mackaness, W., & Webber, B. (2013). Evaluating a city exploration dialogue system combining question-answering and pedestrian navigation. *ACL 2013: 51st Annual Meeting of the Association* *for Computational Linguistics* pp. 1660-1668). The Association for Computational Linguistics (ACL).

Janarthanam, S., Lemon, O., Liu, X., Bartie, P., Mackaness, W., & Dalmas, T. (2013). A Multithreaded Conversational Interface for Pedestrian Navigation and Question Answering. *SIGDIAL 2013*. Metz.

Janarthanam, S., Lemon, O., Liu, X., Bartie, P., Mackaness, W., Dalmas, T., & Goetze, J. (2012). Integrating location, visibility, and Question-Answering in a spoken dialogue system for Pedestrian City Exploration. *SIGDIAL*. South Korea.

Kelley, J. F. (1983). An empirical methodology for writing user-friendly natural language computer applications. *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* pp. 193-196). ACM.

Lemon, O. (2012). Conversational interfaces. *Data-Driven Methods for Adaptive Spoken Dialogue Systems* pp. 1-4). Springer.

Lemon, O., & Gruenstein, A. (2004). Multithreaded context for robust conversational interfaces: Context-sensitive speech recognition and interpretation of corrective fragments. *ACM Transactions on Computer-Human Interaction (TOCHI), 11*, 241-267.

Li, X., & Roth, D. (2002). Learning question classifiers. *Proceedings of the 19th international conference on Computational linguistics-Volume 1* pp. 1-7). Association for Computational Linguistics.

Liarokapis, F., Mountain, D., Papakonstantinou, S., Brujic-Okretic, V., & Raper, J. (2006). Navigation methods for urban environments: using Virtual and Augmented Reality Interfaces. *International Conference on Computer Graphics Theory and Applications*. Setubal, Portugal: Eurographics.

Long, S., Aust, D., Abowd, G., & Atkeson, C. (1996). Cyberguide: prototyping context-aware mobile applications. *Conference on Human Factors in Computing Systems* pp. 293-294). Vancouver.

Loomis, J. M., Golledge, R. G., & Klatzky, R. L. (1998). Navigation system for the blind: Auditory display modes and guidance. *Presence-Teleoperators and Virtual Environments, 7*, 193-203.

Mattos (2011). Consumer GPS/GLONASS Accuracy and Availability Trials of a One-Chip Receiver in Obstructed Environments. *GPS World*.

May, A. J., Ross, T., & Bayer, S. H. (2005). Incorporating landmarks in driver navigation system design: An overview of results from the REGIONAL project. *Journal of Navigation, 58*, 47-65.

May, A. J., Ross, T., Bayer, S. H., & Tarkiainen, M. J. (2003). Pedestrian navigation aids: information requirements and design implications. *Personal and Ubiquitous Computing, 7*, 331-338.

McTear, M. F. (2002). Spoken dialogue technology: Enabling the conversational user interface. *Acm Computing Surveys, 34*, 90-169.

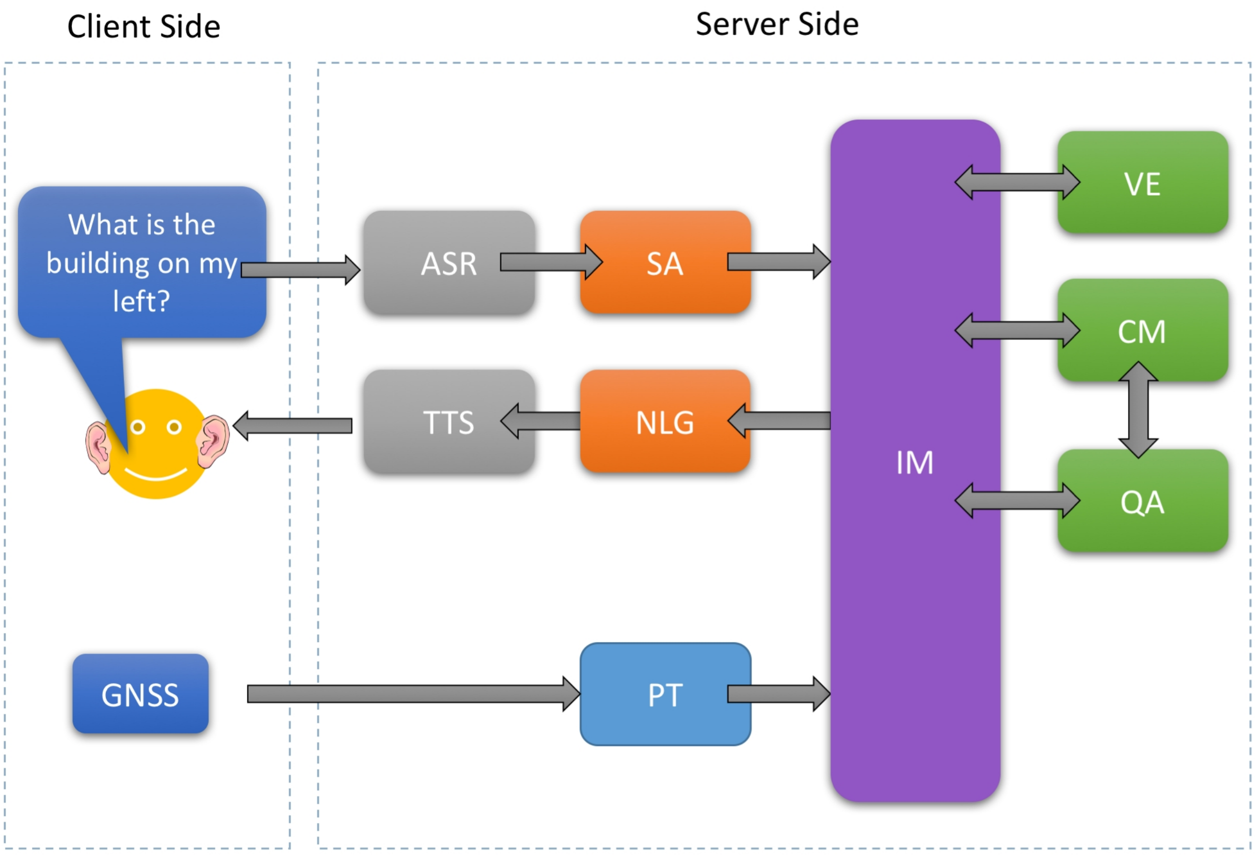
Meek, S., Priestnall, G., Sharples, M., & Goulding, J. (2013). Mobile capture of remote points of interest using line o fsight modelling. *Computers & Geosciences, 52*, 334-344.

Meng Yu, Z. L., Yongqi Chen, and Wu Chen. (2006). Improving Integrity and Reliability of Map Matching Technique. *Journal of Global Positioning Systems, 5*, 40-46.

Mikhailian, A., Dalmas, T., & Pinchuk, R. (2009). Learning foci for question answering over topic maps. *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers* pp. 325-328). Association for Computational Linguistics.

Miller, G. A. (1995). WordNet: A Lexical Database for English. . *Communications of the ACM, 38(11)*, 39-41
.
Montello, D. (1993). Scale and multiple psychologies of space. *Spatial Information Theory A Theoretical Basis for GIS*, 312-321.
Mountain, D., & Raper, J. (2001). Positioning techniques for location-based services (LBS): Characteristics and limitations of proposed solutions. *Aslib Proceedings, 53*, 404-412.
Narzt, W., Pomberger, G., Ferscha, A., Kolb, D., Moller, R., Wieghardt, J., Hortner, H., & Lindinger, C. (2006). Augmented reality navigation systems. *Universal Access in the Information Society*, 1-11.
Raper, J., Gartner, G., Karimi, H., & Rizos, C. (2007). A critical evaluation of location based services and their potential. *Journal of Location Based Services, 1*, 5-45.
Richter, K.-F., & Winter, S. (2014). *Landmarks*: Springer International Publishing.
Saksamudre, S. K., Shrishrimal, P., & Deshmukh, R. (2015). A Review on Different Approaches for Speech Recognition System. *International Journal of Computer Applications, 115*.
Sorrows, M., & Hirtle, S. (1999). The nature of landmarks for real and electronic spaces. In C. Freksa & D. Mark (Eds.), *Spatial information theory* pp. 37-50). Springer.
Strassman, M., & Collier, C. (2004). Case study: Development of the find friend application. *Location-Based Services*, 27–40.
Vlachos, A., & Clark, S. (2014). A new corpus and imitation learning framework for context-dependent semantic parsing. *Transactions of the Association for Computational Linguistics, 2*, 547-559.
Woodsend, K., & Lapata, M. (2011). Wik-isimple: Automatic simplification of wikipedia articles. *AAAI Conference on Articial Intelligence (AAAI)*. California: AAAI Press.
Young, S., Gasic, M., Thomson, B., & Williams, J. D. (2013). POMDP-Based Statistical Spoken Dialog Systems: A Review. *Proceedings of the IEEE, 101*, 1160-1179.


Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

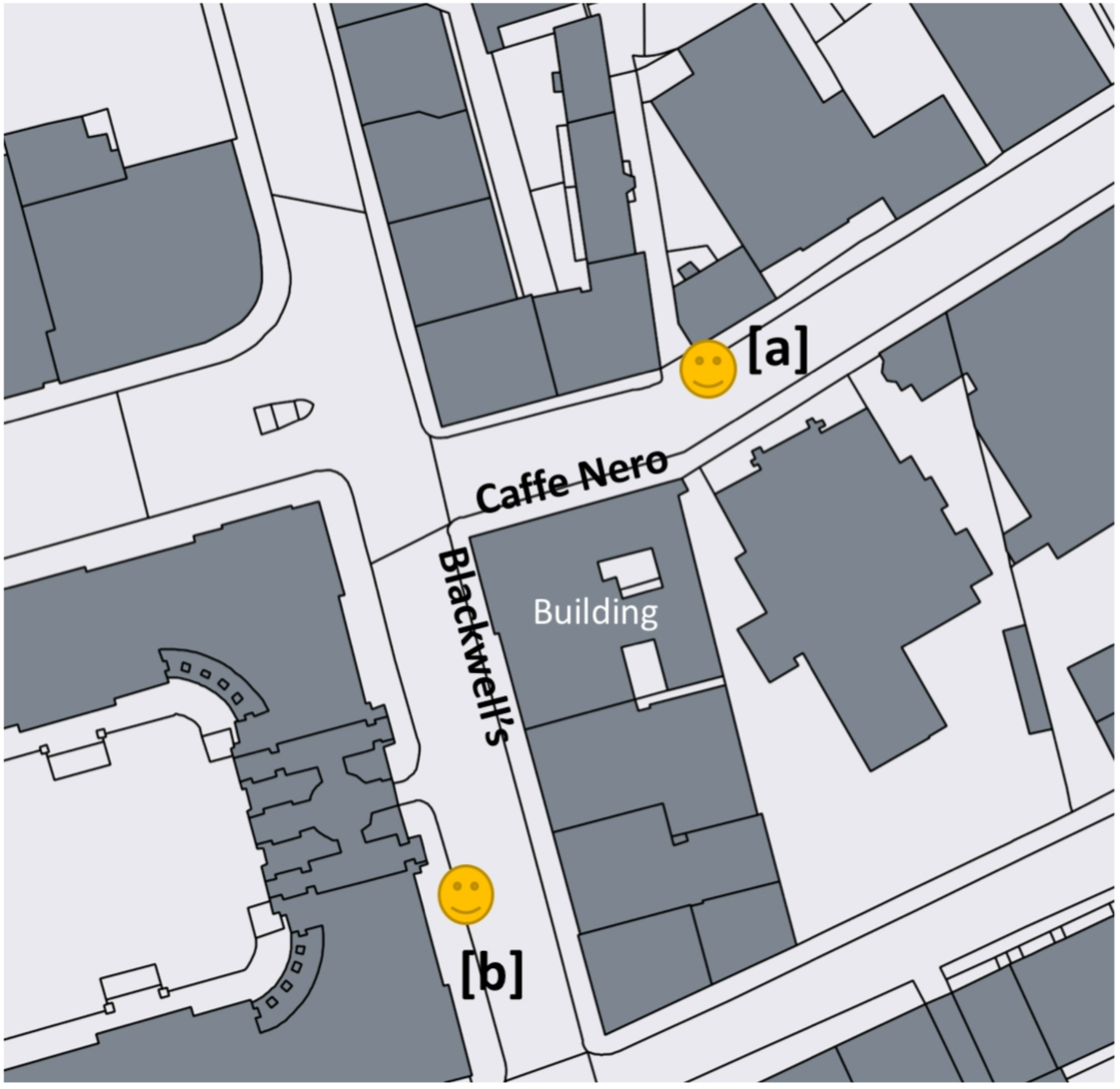


Component List

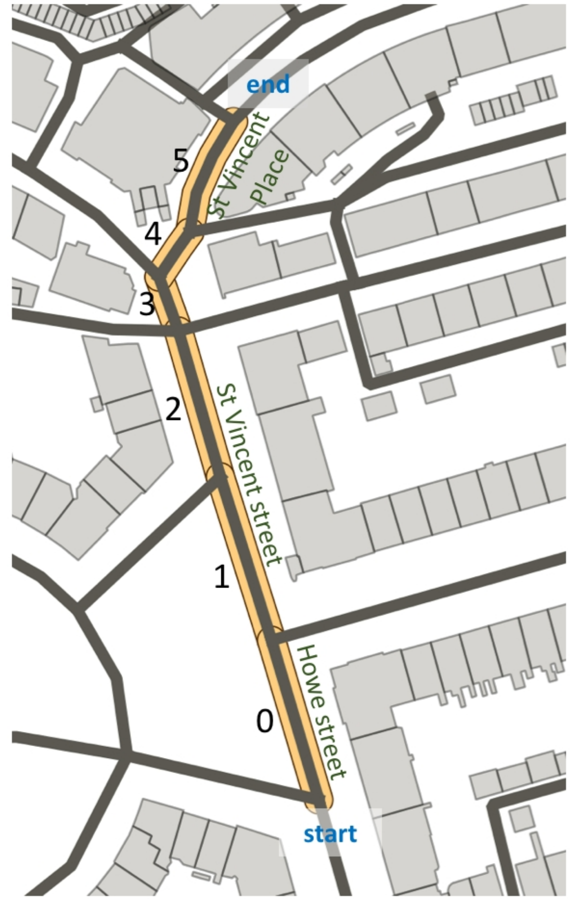
ASR	automatic speech recognition
SA	semantic analysis (natural language understanding)
IM	interaction manager
VE	visibility engine
CM	city model
QA	question answering
GNSS	global navigation satellite system
PT	pedestrian tracker
NLG	natural language generation
TTS	text to speech

USER: what is this church?
dialogAct(set_question)
*isA(id:X2, type:church)
index(id:X2)

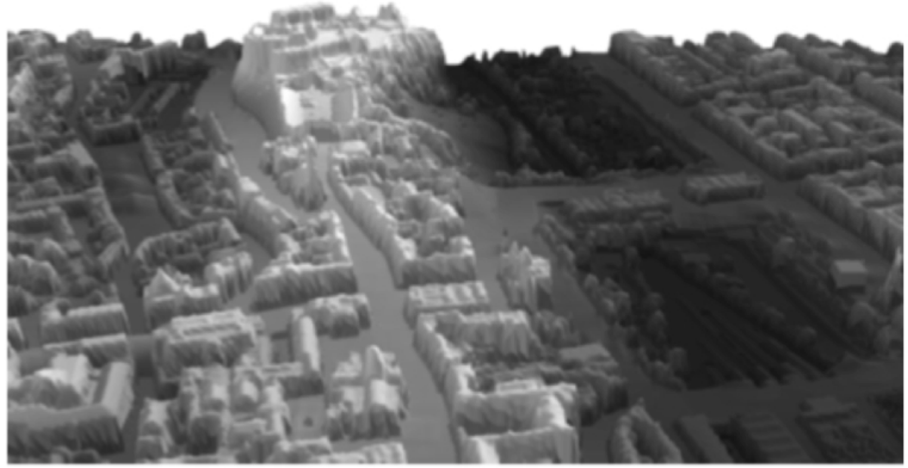
WIZARD: keep walking straight down clerk street
dialogAct(instruct)
*walk(agent:@USER, along_location:X1, direction:forward)
isA(id:X1, type:street)
isNamed(id:X1, name:"clerk street")


[a]
Caffe Nero
Blackwell's
Building
[b]

seq num	edge id	name	cost (m)	path type	bendy	known score	node	next turn	next slope	exits	jtype	exit num	exit desc
0	803879	Howe Street	49.5	street	-	163	2007658	-0.91	downhill	2	-	1	carry on
1	803848	St Vincent Street	51.1	street	-	28	2007610	1.39	downhill	2	-	2	carry on
2	803824	St Vincent Street	45.1	street	-	28	2007568	-3.83	slight downhill	3	X	2	carry on
3	803811	St Vincent Street	17.1	street	-	28	2007547	53.5	flat	2	Y	2	take right fork
4	803812		16.2	street	-		2007581	-25.9	flat	2	Y	1	take left fork
5	803831	St Vincent Place	38.5	street	slight bend	4	2007628	0	null	2	-	-	



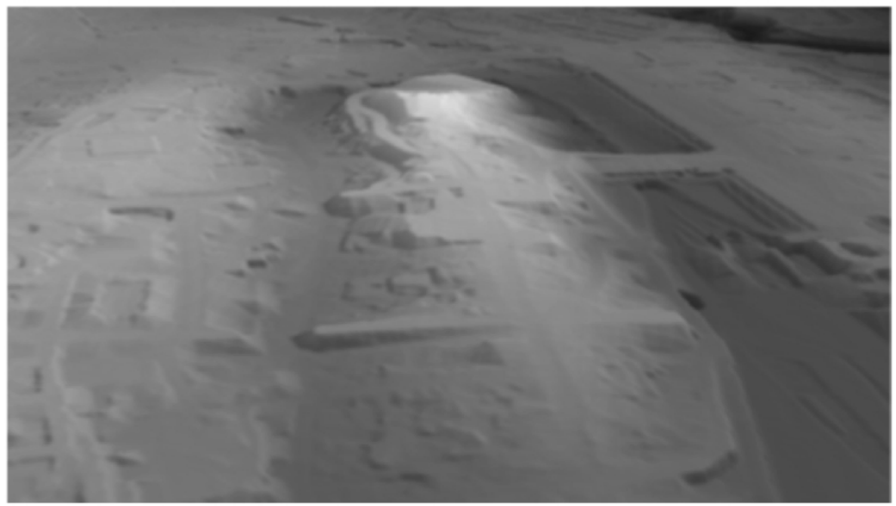
DSM

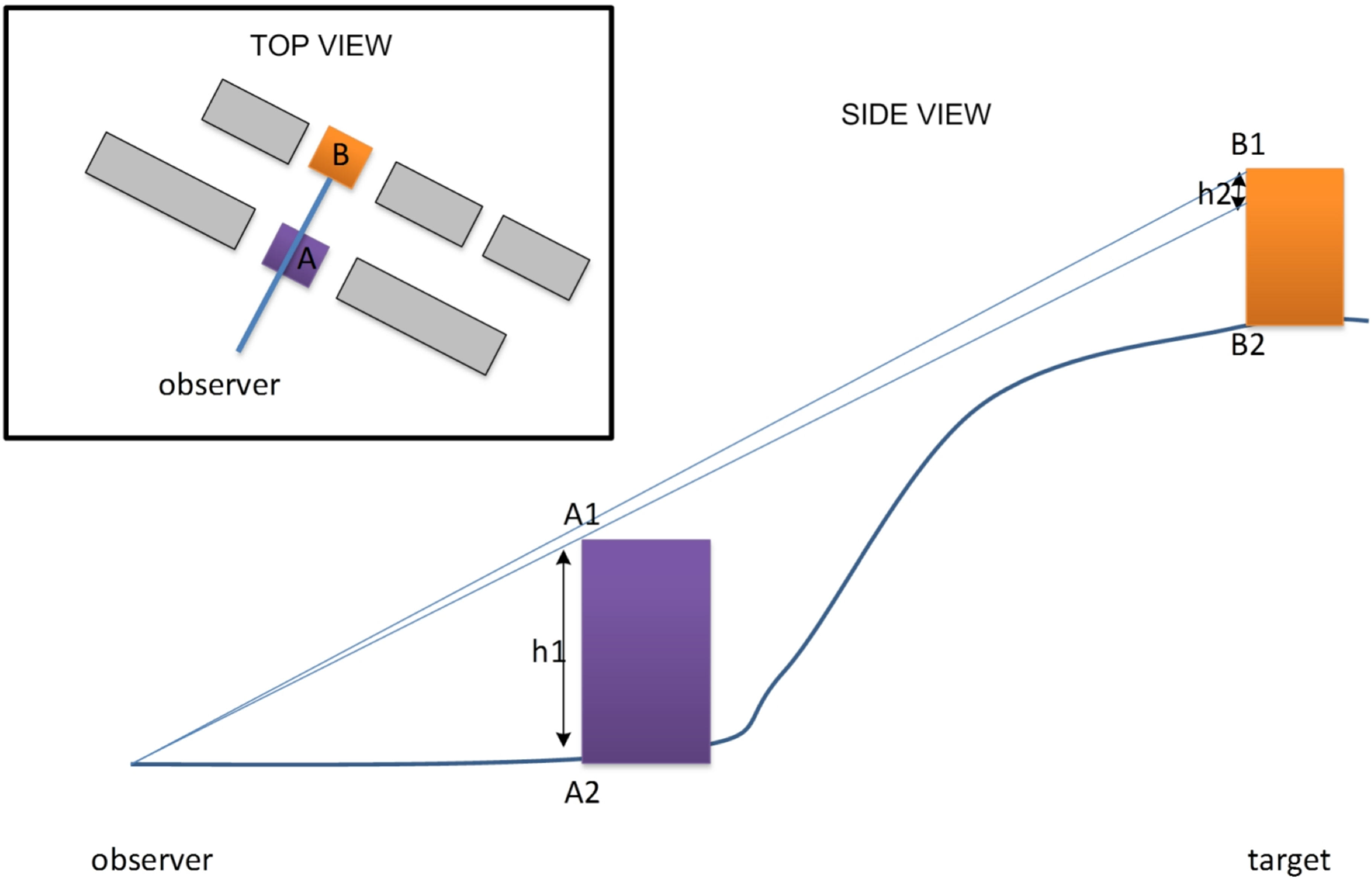


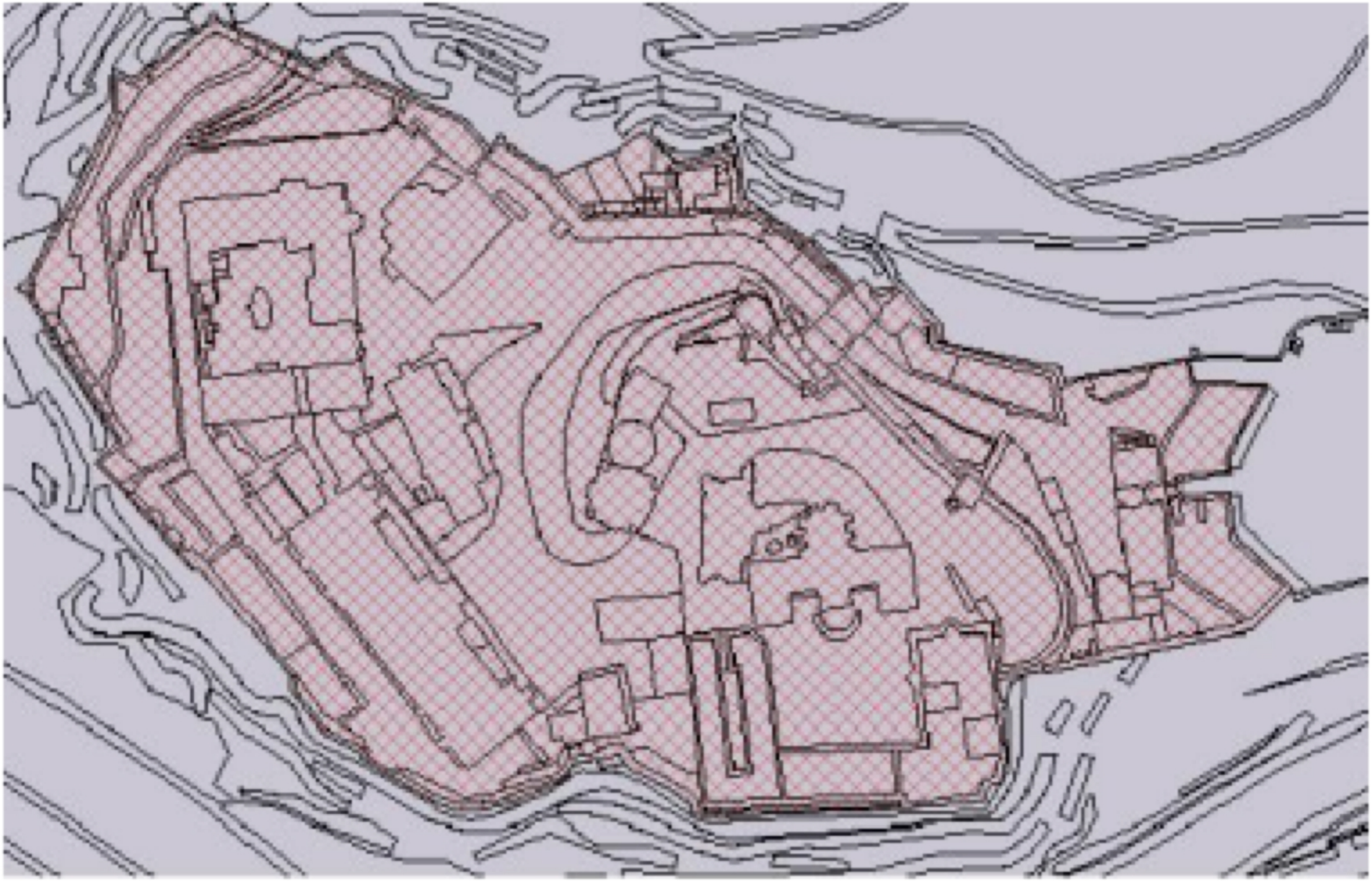
DSM – Perspective View

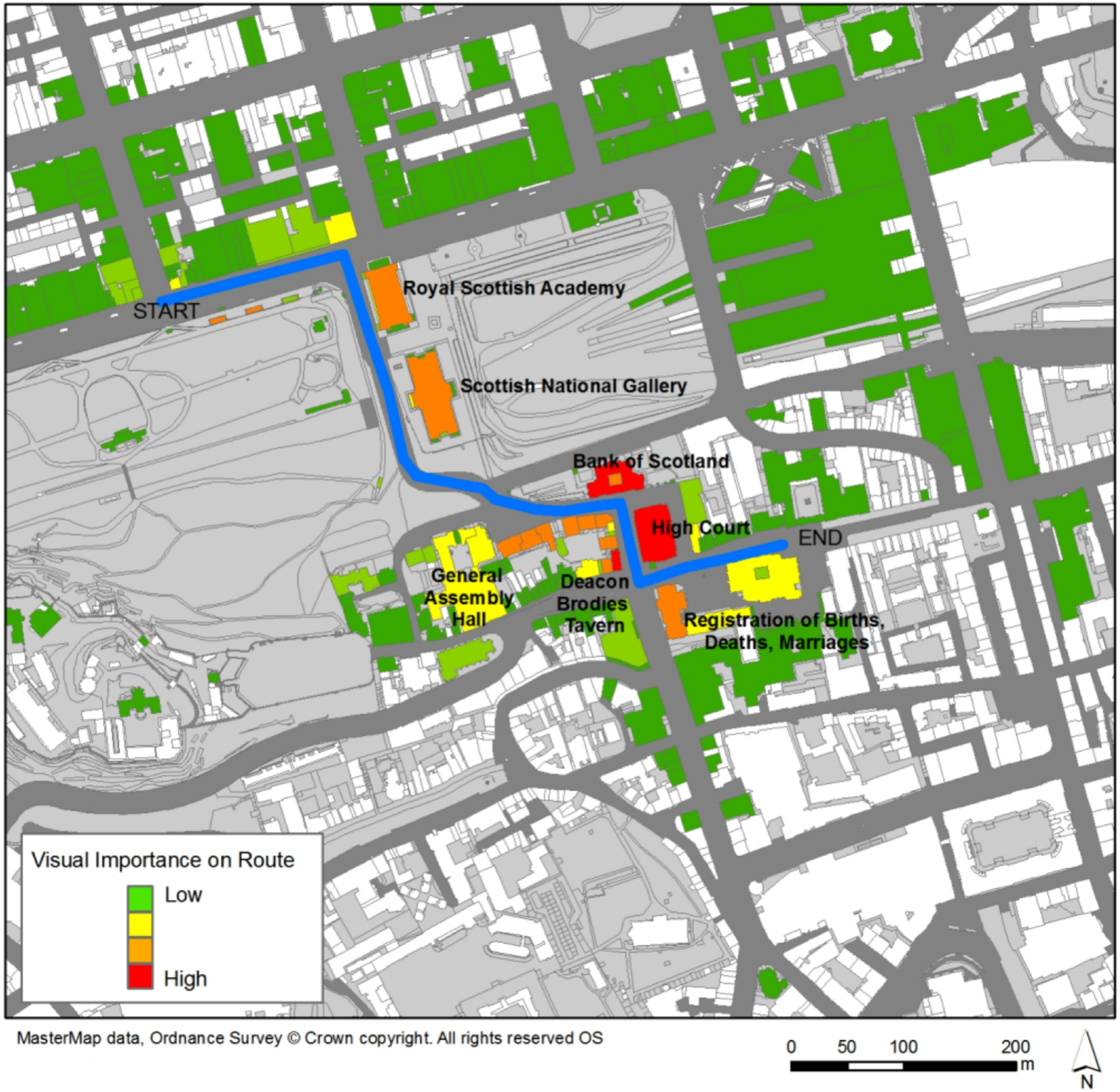
DTM

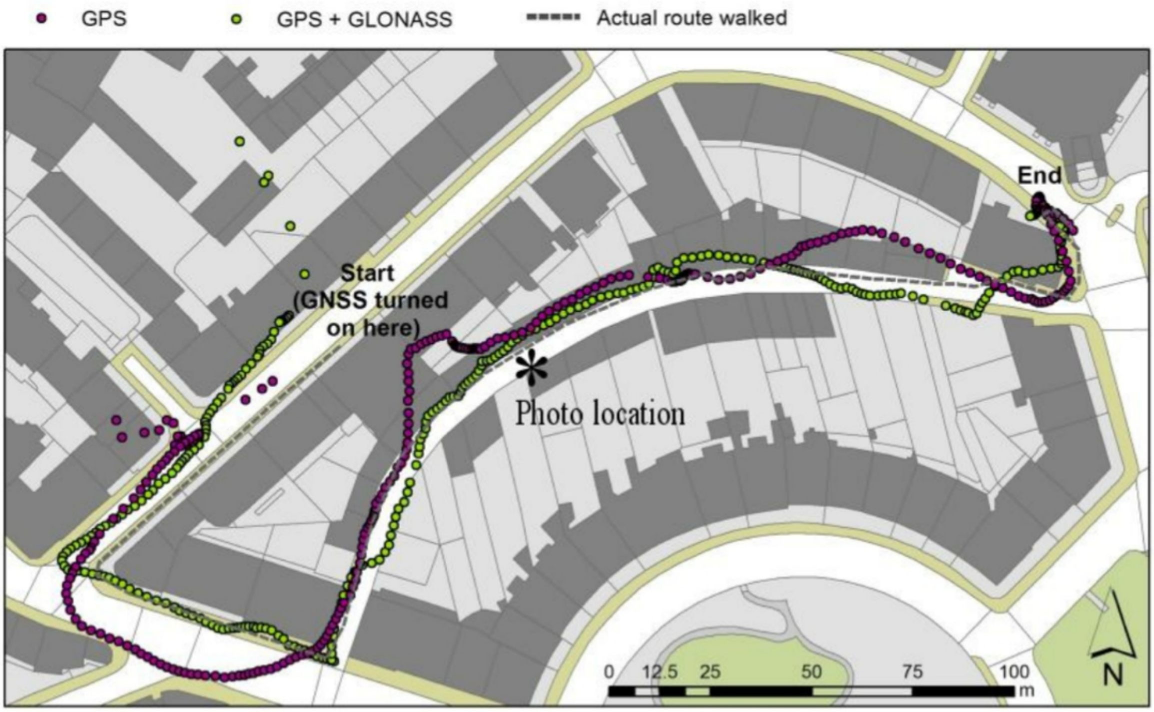
DTM – Perspective View



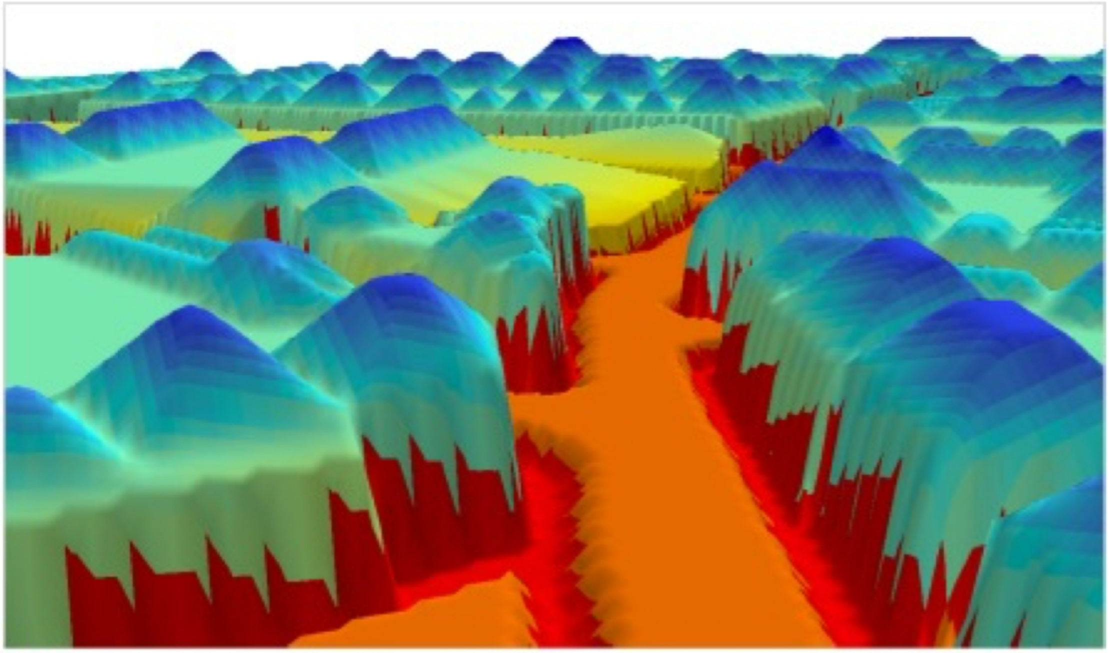

TOP VIEW
B
A
observer
SIDE VIEW
B1
h2
B2
A1
h1
A2
observer
target



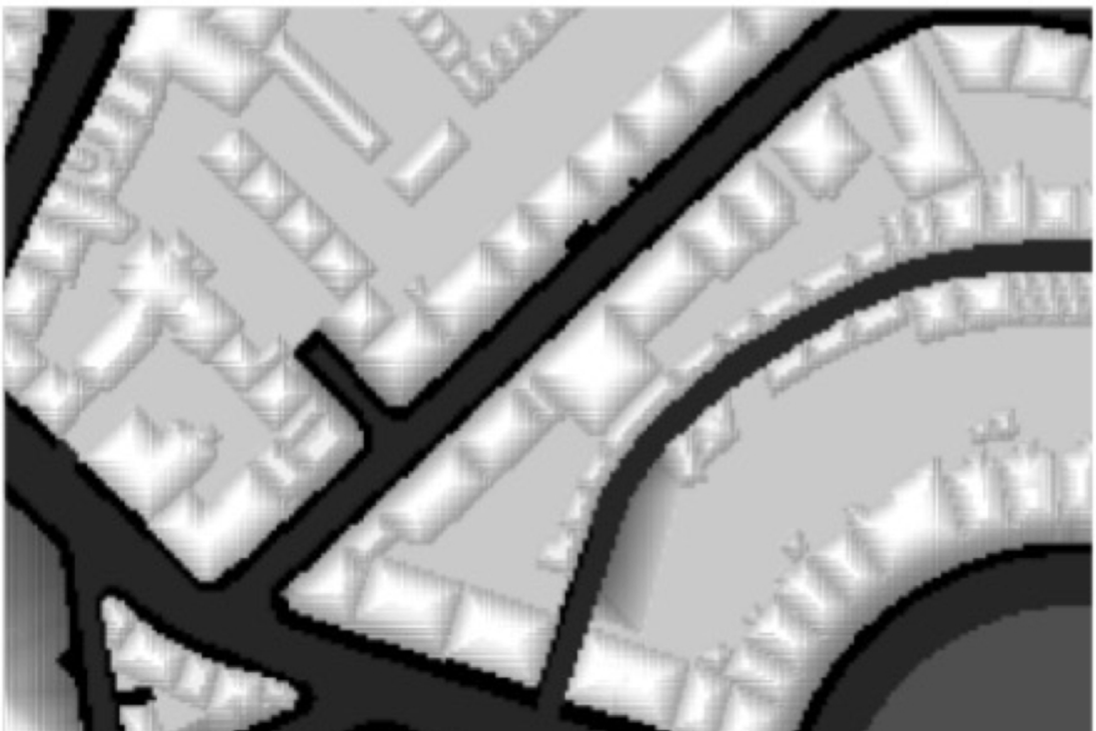

START
Royal Scottish Academy
Scottish National Gallery
Bank of Scotland
High Court
END
General Assembly Hall
Deacon Brodies Tavern
Registration of Births, Deaths, Marriages
Visual Importance on Route
Low
High
0
50
100
200
m
N


GPS
GPS + GLONASS
Actual route walked
Start
(GNSS turned
on here)
End
*
Photo location
0
12.5
25
50
75
100
m
N

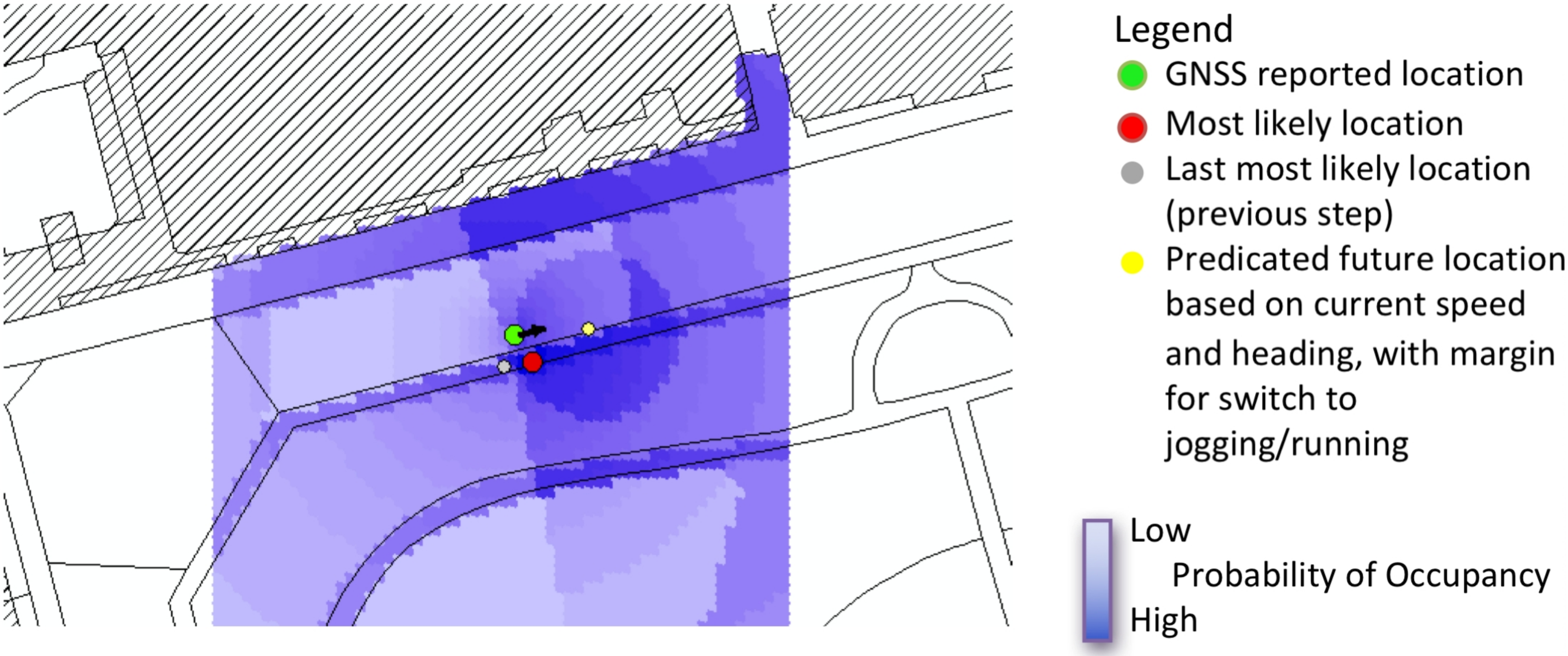
*

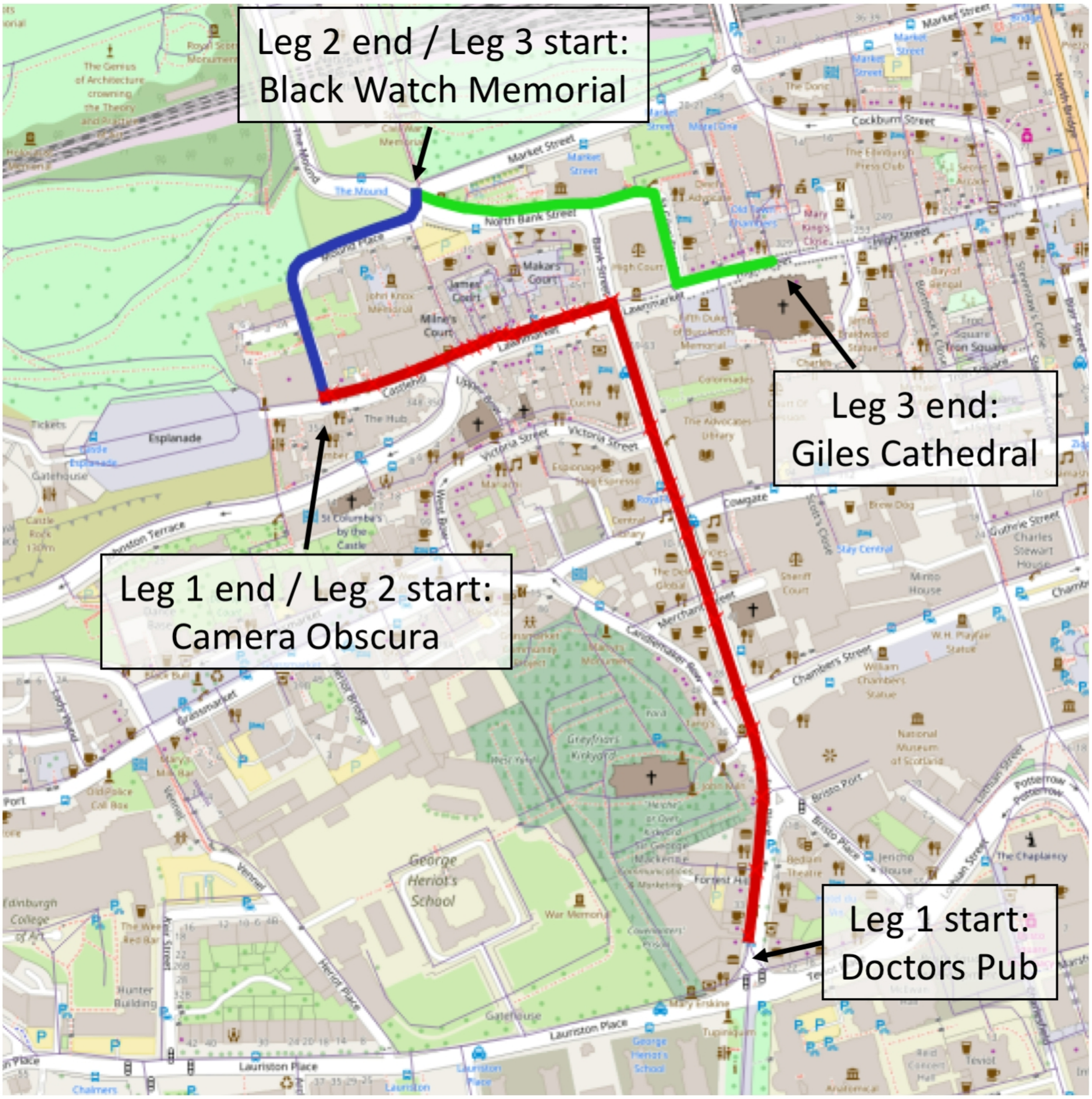


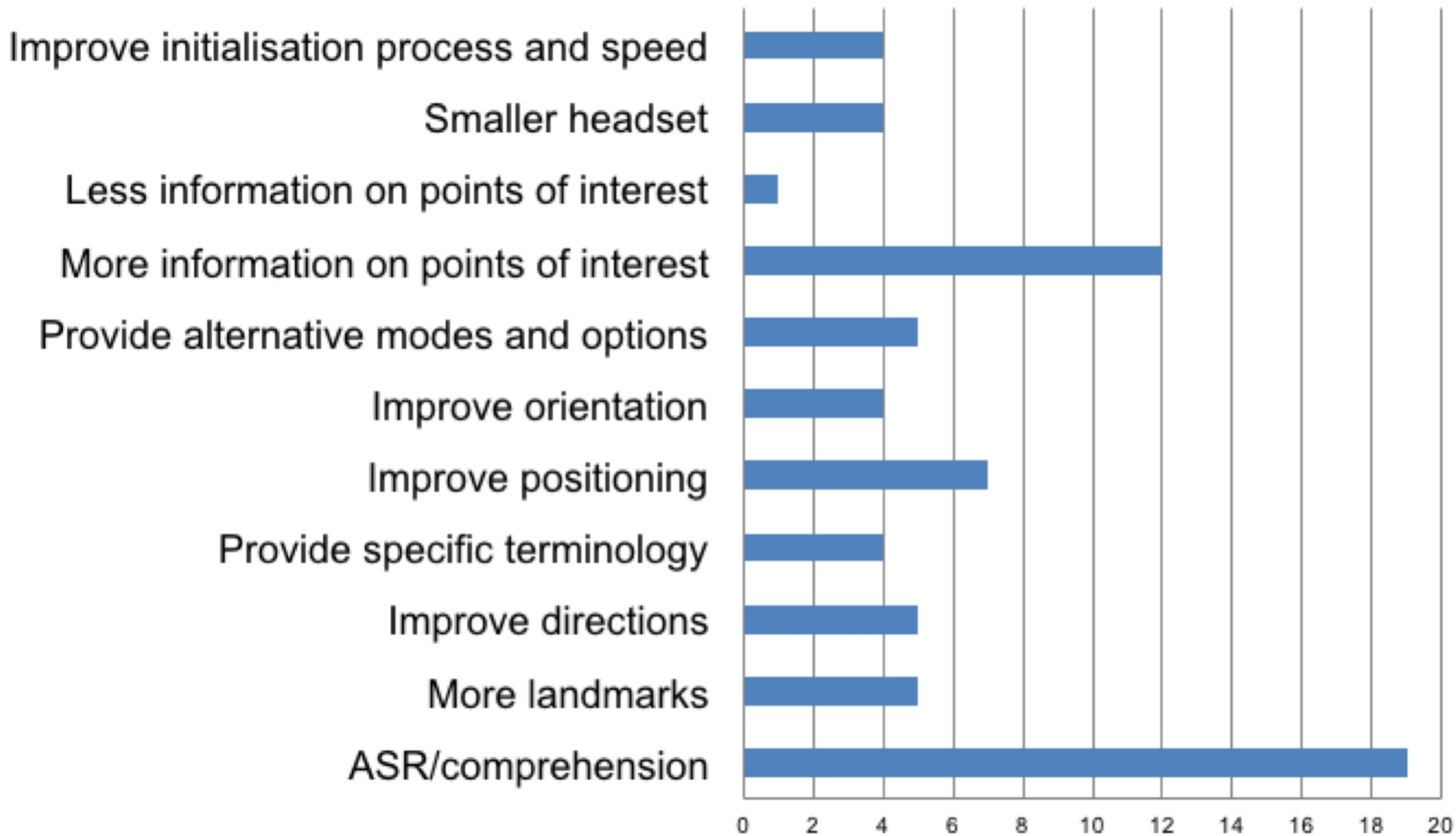
a)



b)


Legend
GNSS reported location
Most likely location
Last most likely location
(previous step)
Predicated future location
based on current speed
and heading, with margin
for switch to
jogging/running
Low
Probability of Occupancy
High


Leg 2 end / Leg 3 start:
Black Watch Memorial
Leg 3 end:
Giles Cathedral
Leg 1 end / Leg 2 start:
Camera Obscura
Leg 1 start:
Doctors Pub


Improve initialisation process and speed
Smaller headset
Less information on points of interest
More information on points of interest
Provide alternative modes and options
Improve orientation
Improve positioning
Provide specific terminology
Improve directions
More landmarks
ASR/comprehension
0
2
4
6
8
10
12
14
16
18
20