Transactions in GIS, 2006, 10(1): 63–86

Research Article

Development of a Speech-Based Augmented Reality System to Support Exploration of Cityscape

Phil J. Bartie
Institute of Geography
University of Edinburgh

William A. Mackaness
Institute of Geography
University of Edinburgh

Abstract

When people explore new environments they often use landmarks as reference points to help navigate and orientate themselves. This research paper examines how spatial datasets can be used to build a system for use in an urban environment which functions as a city guide, announcing Features of Interest (FoI) as they become visible to the user (not just proximal), as the user moves freely around the city. Visibility calculations for the FoIs were pre-calculated based on a digital surface model derived from LIDAR (Light Detection and Ranging) data. The results were stored in a text-based relational database management system (RDBMS) for rapid retrieval. All interaction between the user and the system was via a speech-based interface, allowing the user to record and request further information on any of the announced FoI. A prototype system, called *Edinburgh Augmented Reality System (EARS)*, was designed, implemented and field tested in order to assess the effectiveness of these ideas. The application proved to be an innovative, 'non-invasive' approach to augmenting the user's reality.

1 Introduction

Landmarks support the building of mental representations (Tversky 1993, Hirtle and Heidorn 1993) and are an inevitable means by which people memorize, orientate, wayfind and communicate about space (Lynch 1960, Lovelace et al. 1999, Denis et al. 1999, Michon and Denis 2001, Tom and Denis 2003). This paper reports on research in the field of Location Based Services (LBS) that delivers landmark information via a mobile device. While most LBS provide map viewing on limited displays, this research examines the viability of a device which exclusively uses a speech-based human computer interface

Address for correspondence: William Mackaness, Institute of Geography, The University of Edinburgh, Drummond St, Edinburgh EH8 9XP, UK. E-mail: William.mackaness@ed.ac.uk

(HCI) for imparting geographical information. The non-invasive nature of the interface means the user is less aware of the system, learning about their surroundings just by exploring. This integration of computers into the environment is known as ubiquitous computing and as such, this work fits within the context of Augmented Reality (AR) research (Rauterberg 2002). AR is an exciting area of research offering the ability to enjoy learning about an environment unfamiliar to us by supplementing our physical experience with information held in a computer (Feiner et al. 2004). This paper will examine the methodology, design overview, data creation, design considerations, and evaluation of a prototype system that uses a speech-based interface to provide AR information.

2 Design Aspects

It is relatively straightforward to design an LBS that delivers spatial information in map form, about features as they become physically proximal to the wandering tourist. But activities such as route planning and navigation can be supported by knowledge of all landmarks within the field of view, not just those that are physically proximal. At any given instant, the number of landmarks or features of interest that a tourist can see will change depending on whether they are in open space or 'canyoned' in a narrow street. Two core design issues need to be addressed: the first is how to determine what is in the user's field of view at any given instant, and the second is how to convey information to the user about where and what FoI are in the field of view. In order to know what is in the user's field of view at any given instant, we require: (1) in real time, knowledge of the user's location and facing direction; and (2) a model of the cityscape from which we can determine the degree of visibility of any given landmark. For this we require a spatial database containing a corresponding viewshed for each feature of interest, that can be searched in x, y, z.

The second design issue is to convey this information in a way that is unobtrusive to the user, provides the information in a natural, digestible form, and in a manner that does not make the user feel conspicuous – as they would if they wore a head mounted display for example (Feiner 2002). It was therefore decided that a speech-based interface be built that: (1) was intuitive and easy to learn (Salmon and Slater 1987); and (2) minimised interaction time and supported relatively simple, two way communication. Two way communication required a speech engine capable of text-to-speech (voice synthesis), and speech-to-text (voice recognition) processing – to enable the user to request more detailed information or to record new information. A variety of approaches exist to natural language generation (Reiter and Dale 2000), some of which have been applied to LBS applications – such as route description (Dale et al. 2003). In the context of this project, Table 1 lists the advantages and disadvantages of such a system.

Designing natural language interfaces to descriptions of geographic space requires qualitative descriptions of space to be created from non-linguistic quantitative data sources (Reiter and Dale 2000). Various attempts have been made to formalise people's use of spatial relations in natural language form (Egenhofer and Shariff 1998). For an overview and classification of qualitative approaches to describing geographic space see Frank (1992). A variety of reference frames exist for describing spatial relationships between the FoI and the user (Levinson 1996); in the context of LBS it was appropriate to consider an egocentric perspective, whereby the information was conveyed relative to

Table 1 Advantages and disadvantages of a speech-based, tourist city guide

Advantages
Low power consumption compared to LCD
Natural conversational communication
No distraction from viewing the surroundings (hands free, eyes free)
Accessible to visually impaired people
Lightweight hardware (headphones, microphone), inexpensive – unlike head mounted displays
Compact, yet without the constraints of limited screen area and map design
Secure and discreet – the user may not want to be seen looking at maps or appearing lost
Disadvantages
Speech recognition errors in noisy streets
User's accent and speed of speaking affects accuracy of voice recognition (system coaching required)
Does not allow a user to browse the information
Can not be used by hearing impaired

the user's location and facing direction (i.e. a relative frame of reference). Finally, in order to moderate the flow of announcements, a methodology was required to calculate the significance of FoI according to factors such as their degree of visibility and size, their angle relative to the facing direction of the user, and the time elapsed since they were announced.

3 Methodology

The research was undertaken in five stages: the first stage involved selecting a city and associated FoI (key landmarks and tourist sites). The second phase (section 4) was to create a digital surface model (DSM) of that city (this was achieved using LIDAR data). To support real time use, the database had to be configured in such a way that it supported very high speed retrieval of information on the FoI and its degree of visibility. The third stage was to use locational technology to determine the user's location and facing direction (this was achieved using GPS technology). This required field testing and development of techniques that dealt with problems such as poor signal strength. The fourth component (section 5) related to the creation of natural language descriptions of what and where each FoI was, in the user's of field of view. The descriptions utilised a text-based gazetteer. The fifth component (section 6) was an interface that supported two way interaction via a restricted vocabulary. All of these elements were brought together in a prototype called Edinburgh Augmented Reality System (*EARS*) (Section 7), for evaluation in both field and user tests (section 8). *EARS* was modified by iterative design throughout the life span of the project. These five stages are discussed below in more detail, and summarised in Figure 1.

3.1 Features of Interest

Edinburgh city in Scotland was chosen as the test region because it was known to the authors, and has many widely distributed FoI. It offers many different urban landscapes

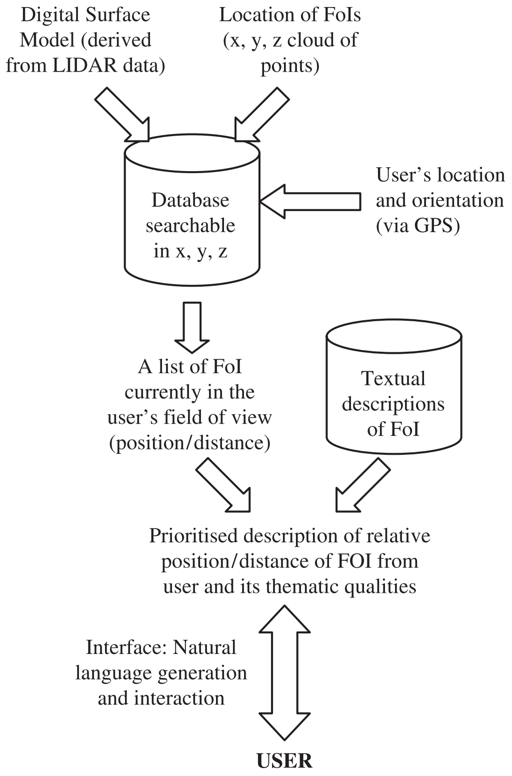

Digital Surface Model (derived from LIDAR data)
Location of FoIs (x, y, z cloud of points)
Database searchable in x, y, z
User's location and orientation (via GPS)
A list of FoI currently in the user's field of view (position/distance)
Textual descriptions of FoI
Prioritised description of relative position/distance of FOI from user and its thematic qualities
Interface: Natural language generation and interaction
USER

Figure 1 System components

arranged on a multi-level street system making navigation and orientation often tricky for newcomers. If a system design could function well in Edinburgh then it ought to be adaptable to many urban areas. Eighty six FoIs were chosen from around the city, varying in shape and height, including features of interest to tourists such as Edinburgh Castle, Tron Kirk Information Centre, Holyrood House as well as monuments and statues. The FoI were chosen so as to be dispersed throughout the city, and ranging in size so visual acuity could be examined. The selection needed to be numerous so that issues of announcement queue bottle necking and significance ordering could be examined (discussed later in this paper).

4 Creating the Viewshed

The second stage was to compute the degree of visibility for any given FoI for any position in the cityscape. The viewshed is the resultant map produced after performing a visibility analysis from observation points, based on a Digital Surface Model (DSM). Building and vegetation heights are required to calculate urban viewsheds, therefore LIDAR (Light Detection and Ranging) data – collected by timing the return of a laser beam bounced from an aircraft to the ground – was used (Palmer and Shan 2002). This is a technique that has been used to great effect in determining building heights in urban contexts (Rottensteiner and Briese 2002). Two 2 km by 2 km tiles of LIDAR data with a ground resolution of 1 m by 1 m were sourced from the Environment Agency. As it was intended that any user of the system could enjoy free exploration of the city without being guided by the system, it was necessary to work out from which parts of the city each FoI could be viewed. This could be done by either selecting each LIDAR cell in turn and calculating what was visible from that location, or alternatively by working out the areas of the city from which each FOI could be seen. In either approach the degree of visibility is also dependent on the height of the user. For this project a height of 170 cm was assumed.

Of the two approaches, the first approach would require each of the 8 million LIDAR cells to have a viewshed calculated and stored, something which has been attempted before and is called the Complete Intervisibilty Database (CID) (Mineter et al. 2004). A small test showed that the calculation time for the selected pilot area would take many months. Since it was not practical to consider every cell for the chosen study area, the observation points were sub-sampled and placed every 10 m and only on the street level surfaces (i.e. it was assumed the user would only ever travel the streets and so no views from the rooftops were calculated). This resulted in only 800 viewsheds being required. Furthermore, since *EARS* only required information on which of the 86 FoI were visible it was possible to compact the data considerably by assessing each FoI's visibility from every road observation point. This process resulted in a dataset of 10,000 records for only a small test area. This was still not sufficiently fast and an alternative approach was considered.

An alternative approach to this complete intervisibility database was to consider which areas of the city would be visible from each of the 86 FoI. Again the user height must be considered; however, careful thought was required as to how each FoI was represented in the database. Rather than a single point, it was necessary to define a cloud of target points for each FoI – some at the base, in the middle and at the top. Far away and the user might only see the top of a FoI. Close to, and the user's view of the top of the FoI may be obscured by its walls. Targets are therefore required at the base in order to prevent the system concluding that the FoI is out of view. A number of different methods were considered as to how best to represent a building FoI using the minimal number of target points, in an effort to reduce processing time. These included randomly assigning points, manual placement, complete coverage of the FOI, allocation based on the highest point, the FOI centroid, and at all nodes on the outside edge of the building polygon.

The target points chosen needed to reflect the most significant parts of the building shape and also cover approaches from any angle. Evaluations were carried out using a complex building shape (Scots Monument in Edinburgh City Centre). It was found that the best performance came from the manual placement of points. However in some instances neither the boundary, nor the highest point of an FoI might be visible, yet the centre of the FOI could be. This could arise when viewing a large structure such as Waverley Railway Station. Therefore as well as the manually placed points, a second dataset covering the two highest points and the centroid of each FOI was included. The number of manual points placed per FOI varied depending on the complexity of the shape of the feature. For example Edinburgh Castle was covered with 92 target points, while only 20 were used for Saint Giles Cathedral.

The viewsheds were calculated for every point placed, and resulted in a dataset for each FoI that recorded the frequency of targets which could be viewed from each square metre of the city (taking into account the height of the user by applying an offset of 170 cm to each cell). As the number of targets used to represent each FoI varied, the resultant viewshed values were not directly comparable between FoI. Therefore the viewshed data was reclassified to a common scale of 0 to 9 for the manually seeded targets (Dataset A), and 0 to 3 for the maximum and centroid viewsheds (Dataset B). The two datasets were combined – using the function (10A + B), in order to reduce storage into a single viewshed per FoI. The system was designed for street level use only, and therefore to reduce the dataset size still further a rasterized Ordnance Survey (OS) MasterMap building layer was used to mask out (Tomlin 1990) the viewshed results from within buildings in the city. Examination of the resulting viewsheds indicated minor referencing and resolution issues between the LIDAR and OS MasterMap datasets had left data fragments along building boundaries. This required cleaning otherwise any GPS locational errors might result in the system announcing a FoI as being visible, while at the street level this might not be the case. The dataset was then exported as a series of ESRI ASCII GRID files and by use of a custom JAVA utility converted into an 'x, y, z' format for loading into a text database system where 'z' stored the degree of visibility, or VRANK. The final step in this phase was to standardise the dataset into a single coordinate system and projection (WGS84) to match the output from the GPS. This entire process is summarised in Figure 2.

The data generated from the viewshed calculation were stored in a local mySQL RDBMS holding 12 million viewshed records. Database tuning (mySQL 2004) was necessary to search this number of records with acceptable system performance, and resulted in first time searches (i.e. not cached results) on a specified WGS84 latitude and longitude being reduced from around 15 seconds to sub-second (typically 0.3–0.6 seconds). Each search result was attributed with the current distance between the user and the FoI, and both the relative and absolute angle. The relative angle takes into account which direction the user is facing, and is required so that *EARS* can notify the user to look left or right to see a new FoI. The absolute angle is used by *EARS* to determine which side of FoI is facing the user, and therefore any announcement information can be customised to highlight specific features visible from the user's location and route of approach.

4.2 Location and Orientation

At any given instant it was critical to know the user's location and facing direction. There are a number of methods which can be used to determine a user's location. Some techniques lend themselves to indoor usage where the environment can be controlled, while others are more suitable outside. A WAAS enabled GPS, with altimeter and digital compass, was chosen as the method for locating the user in this research as it did not require any installation or maintenance on the part of the system designer, its use was free, and previous work on a *Mile Mapper* (Revell 2001) application had shown that it provided an adequate solution when walking among the old parts of Edinburgh City. It is acknowledged that GPS does not provide a perfect location solution for an AR device (Feiner 2002), in particular issues regarding accuracy and signal strength in the urban corridors created between high buildings (Kleusberg and Langley 1990). This is a discussion point in the Evaluation section of this paper.

It was intended that the GPS's digital flux gate compass would provide user heading values as it functions at all times even while the user is stationary, unlike GPS direction which was calculated from the vector derived from the user's movement. However, it was found that even a slight change to the horizontal inclination of the GPS device could alter the output value by up to 180 degrees. Instead the heading was calculated from

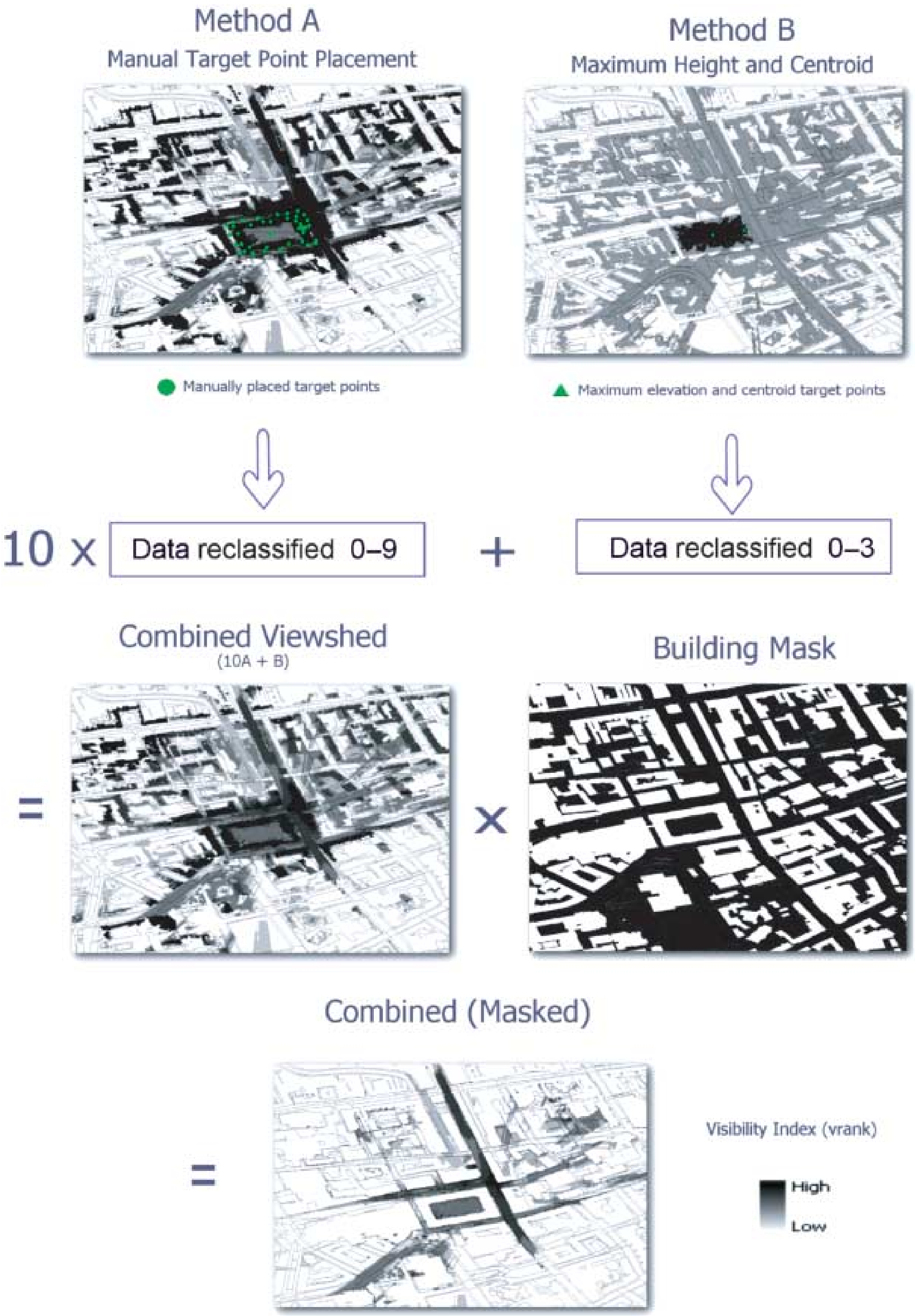


Figure 2 Overview of viewshed dataset creation (LIDAR data copyright Environment Agency, MasterMap data, Ordnance Survey © Crown copyright. All rights reserved OS)
This figure also appears in colour in the electronic version of this article and in the plate section at the back of the printed journal

Table 2 Example sentence construction

Audiotag		Visibility	Direction		Distance
Edinburgh Castle		Just Visible	In front of you		800 metres
Scott Monument	is	Clearly Visible	In front of you just to your right	at	1.1 kilometres
Bedlam Theatre		–	To your left		120 metres

the vector (movement of the user) though this did not accommodate the user who stopped and turned on the spot!

5 Natural Language Generation Describing Location of FoI

The next stage is to take this quantitative information and through natural language generation turn it into a text that can be narrated to the user. Describing spatial data via an auditory interface requires consideration of the sequencing of material (Weber 1998) and the broader issue of communicative acts (Maybury 1993). To inform a user where a FoI was situated *EARS* could report an exact distance and angle, but most users would be unable to quickly decode this into meaningful information. Therefore a general direction and rounded distance were used – converting the relative angle into a generalised qualitative statement. The initial assignment of angle to qualitative statement was derived and revised through field testing. Figure 3 shows the words used according to the calculated angle between the user and the FoI. This, together with attribution from the viewshed database (distance and degree of visibility) enabled the creation of constructs using phrases such as "in front of you", and "just visible". Some examples are given in Table 2.

5.1 *Filtering*

There was a need to prioritise the announcements based on their distance, angle, visibility and size. To ensure that the user did not suffer from information overload (Pashler 1995) from hearing repetitive or irrelevant information, the system filters the results from each search. It was unlikely that a user would wish to hear announcements for a FoI many times in succession just because the object remained visible in the field of view. Similarly a user may enter a courtyard, and at some point retrace their footsteps, but may not wish to hear recent information on FoI repeated. Therefore *EARS* recorded the last announcement time for any FoI and depending on the user's preferences, filtered out any repeat announcements. A number of other optional and customisable filters were included in the prototype, each drawing on attribution drawn from (1) the viewer's position relative to FoI in the database, and (2) their trajectory (drawn from positional and cardinal information gathered in real time as the user moves through the real world). These are summarised in Table 3, and explained in greater detail in Bartie (2004).

5.2 *Weighting*

EARS attempts to judge the importance of each announcement based on a number of factors, specifically the relationship of the user's location with respect to the FoI in the

Table 3 Filters available in EARS

System Filters		
In Front	Only Announce Items which are in front of the user (relative angle from user direction to FoI was between –90 and +90 degrees)	
Far Away	Only announce items which are at great distance if the visibility index was high (i.e. do not announce distant objects if they are barely visible)	
Barely Visible	Do not announce small items (based on the footprint area) if greater than 100 m	
Recent	Do not announce a FoI if it has been announced to the user within the last 2 minutes	Option to ignore this if the absolute angle to the FoI has changed by more than 90 degrees since last announcement (i.e. looking at a different face of the FoI)
	Do not announce a FoI if the user's current location was within 250 m of where the FoI has been announced before (the announcement was close spatially to last announcement location)	
According to Type	Only announce items from specific categories of FoI (such as tourist attractions, University of Edinburgh buildings, government buildings.)	

database in terms of distance, relative angle, size, and visibility. Near items were assigned a high priority (Tobler 1970) so that they may be announced before being passed by. A dynamic queue was generated of the FoI that would be announced to the user. The weightings were recalculated after each announcement and the queue was re-ordered to ensure the system reflected the user's movement through the cityscape.

5.3 Distance

The distance weighting was allocated according to a decay function (equation 1), which was derived empirically through field trials.

$$\text{DistanceWeighting} = 500*(1/(5 + (\text{Log}(\text{dist})\text{\textasciicircum}2))) \qquad (1)$$

An exponential decay curve ensures great importance was attached to FoI nearby whilst FoI further away (say 200 m) were given much less weight, making them less likely to be announced. An override ensures that if the FoI was within 5 m then the item was heavily weighted forcing it to the top of the announcement queue.

5.4 Angle

EARS also weighted the significance of an item depending on its angle relative to the user, by default filtering out items behind the user. Items in front of the user were considered more important than those to the sides and were therefore given a higher weighting factor. Figure 3 shows the weightings used by *EARS* along with the relevant speech phraseology used during the announcements. The phrasing and weighting were refined during field trials.

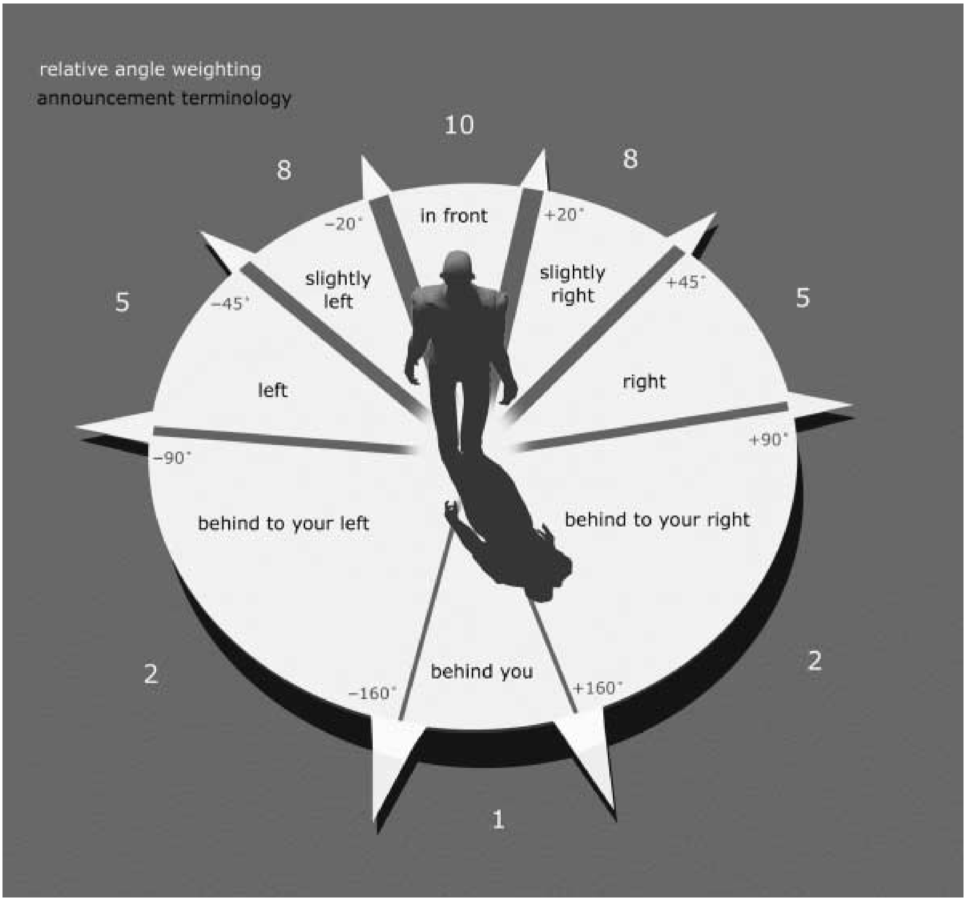


Figure 3 Direction weightings and qualitative phrasing

5.5 Time

The temporal element was important in weighting and filtering announcements. By default a FoI was filtered out from the speech queue if it had previously been announced in the previous 2 minutes. However the user was able to customise or disable this filter, or request *EARS* to bypass the filter on the condition that the approach to the FoI be significantly different from the last time it was announced. This allowed the system to notify the user when a different face of a FoI was available (Bartie 2004). Considering the priority of any items in the announcement queue, a FoI which has been recently announced was less important than one which the user had yet to be made aware of. Therefore the system allocated greater weight to items which had not been announced, or that were announced a long time ago.

The overall weighting value assigned to each FoI in the announcement queue was calculated by taking the product of the amount of visibility, the size of the footprint area of the FoI, its distance and angle relative to the user, and the time since it was last announced. After each announcement the weights were recalculated,

Table 4 Voice commands available in *EARS*

Command	Action
HELP	Notify user of all command words
AUDIOTAGS	List all current FoI with active grammar
SHUT UP	Stop talking and clear announcement queue
SKIP ITEM	Jump to the next item in the announcement queue
ADDRESS	Notify user of the closest listed shop address
CLOSEST POINT OF INTEREST	Notify user of closest feature of interest, direction and distance
ADD WAYPOINT	Add a waypoint to the spatial database, including comment if desired
STATUS	Reports current GPS horizontal accuracy, elevation, LIDAR cell value, digital compass heading, and GPS heading
DISABLE	Stops voice recognition but announcements still occur (only 'pay attention' command recognised)
PAY ATTENTION	Activates voice recognition

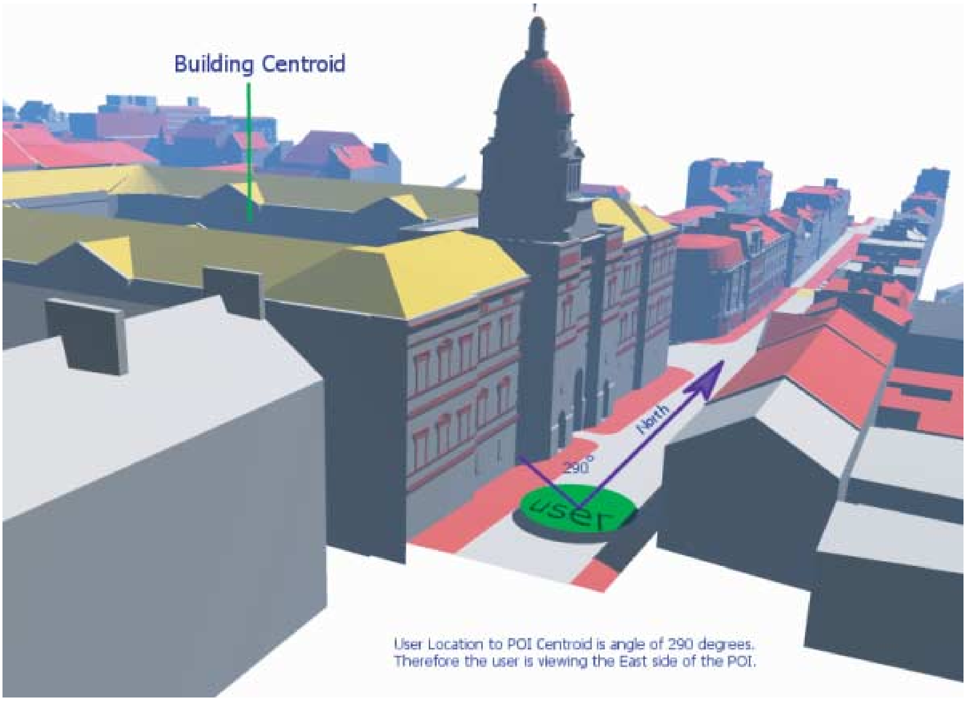
so that any changes in the user's direction, or location could be used to re-adjust the announcement order.

6 Voice Based Interrogation

The user was able to control the system through a set of voice activated commands. These included asking the system to clear the announcement queue, skip over a single item in the queue, announce the closest point of interest or address, obtain help on using the system, or add a new waypoint to the spatial database. There were also commands to check the GPS accuracy intended for system performance testing, and commands to activate or disable the voice recognition functions (Table 4).

The system divides the acceptable grammar into those items always available to the user such as 'help' or 'closest point of interest', and those which become available when appropriate. This approach reduces the number of concurrent words in the active grammar list and thus increases the recognition accuracy and speed of interaction. By 'appropriate' we refer to the idea of FoI becoming part of the acceptable grammar, during and for a while after they become visible to the user. Each FoI has an associated audiotag which was used to announce that the item was visible from the current user location. An alternative recognition tag was also associated with the FoI. This dataset, held in a mySQL table, was loaded on application launch and turned into a grammar list. Each item's recognition tag was made into a grammar rule and set as being 'inactive' (Microsoft 2003). Once the FoI has been announced the rule was made active for a period of three minutes. During this time the system will recognise the FoI if it is spoken by the user, and offer the user detailed information about the FoI should the user request it (the user simply stating the name of the FoI).

EARS checked the audio input for matches against the active grammar rules, with any matches within confidence limits (discussed in Section 8), then triggered the appropriate Visual Basic method.



User Location to POI Centroid is angle of 290 degrees.
Therefore the user is viewing the East side of the POI.

Figure 4 Using XML bookmarks to customise detailed text based on user's absolute angle to POI (3D CAD data copyright EWHT)
This figure appears in colour in the electronic version of this article and in the plate section at the back of the printed journal

6.1 Customised Detailed Text

Functionality within the system allowed the response to a user's request for detailed information on a FoI to be customised based on their angle of approach – the benefit being that the system was able to notify the user of the most relevant details of a structure on the closest face. This was made possible by the inclusion of XML tags within the detailed text field, known as bookmarks. These bookmarks were not turned into voice announcements by the speech handler but processed according to their content. The speech handler recognised bookmarks which signified relevant text to be announced when the user approached from a particular angle, or was within close distance of a feature. These bookmarks were optional, but could be used in combinations so that a single FoI might have many different detailed texts according to the user's view of the feature. For example when approaching the Institute of Geography from the south the system would include information on the historic gate posts (only viewable from this approach), but when approaching from the west it would inform the user of the car park leading to the Drummond Library. Currently bookmarks exist for approaching a FoI from the north, east, south, west, or when the user was within 100 m, but it would be very easy to add bookmarks for customised distances or approach angles.

For example in Figure 4 a user requesting full details on the Old College while looking west (i.e. viewing the east side of the building) will be notified of the gold figure

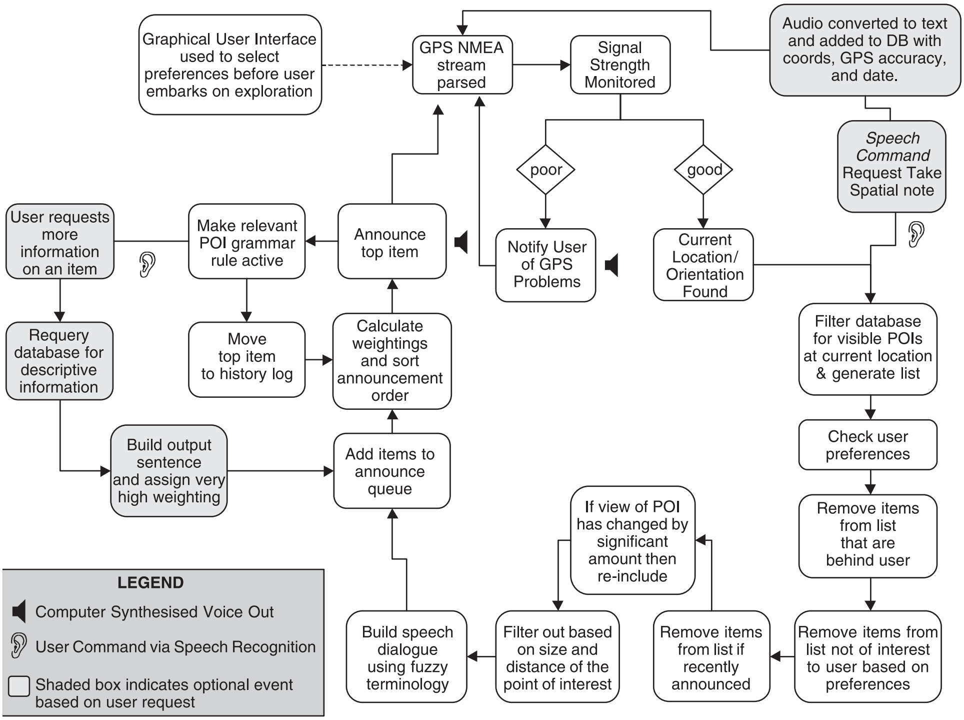
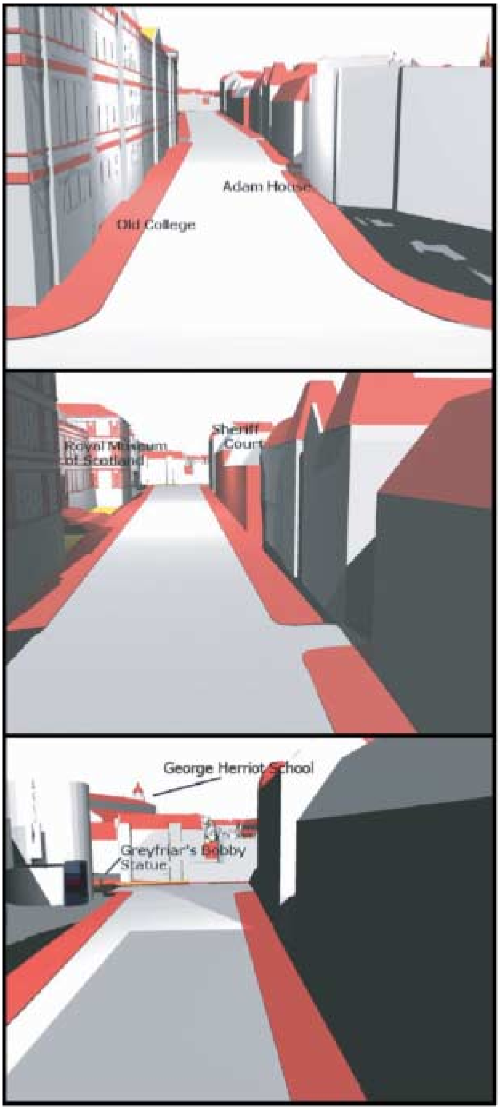

Graphical User Interface used to select preferences before user embarks on exploration
GPS NMEA stream parsed
Signal Strength Monitored
Audio converted to text and added to DB with coords, GPS accuracy, and date.
Speech Command Request Take Spatial note
poor
good
Notify User of GPS Problems
Current Location/ Orientation Found
User requests more information on an item
Make relevant POI grammar rule active
Announce top item
Requery database for descriptive information
Move top item to history log
Calculate weightings and sort announcement order
Filter database for visible POIs at current location & generate list
Check user preferences
Build output sentence and assign very high weighting
Add items to announce queue
If view of POI has changed by significant amount then re-include
Remove items from list that are behind user
Build speech dialogue using fuzzy terminology
Filter out based on size and distance of the point of interest
Remove items from list if recently announced
Remove items from list not of interest to user based on preferences
LEGEND
Computer Synthesised Voice Out
User Command via Speech Recognition
Shaded box indicates optional event based on user request

Figure 5 System components and operation

on the dome because it is viewable from this approach. The text that thematically described the FoI itself could come from guidebooks, or any descriptive source. It could also be provided dynamically from internet news feeds. For example 'Really Simple Syndication' (RSS) can be used as a standard way to supply real time data from a remote server to a client such as *EARS* (Pilgrim 2002). In this project, it came from the Gazetteer for Scotland, the first comprehensive gazetteer for Scotland since 1885, it contains details on towns, tourist attractions and historical sites in Scotland (http://www.geo.ed.ac.uk/scotgaz/).

7 Implementation

EARS was written in Visual Basic 6 on a PC laptop platform. The basic sequence of system operation was to locate the user, then search the database for all visible items calculating the distance and angle to each result. A set of filters was then applied to remove items according to the set of predefined rules and user preferences. The remaining items were added to the speech handler's announcement queue and assigned a weighting value, the queue was then sorted and announced to the user. The announcement queue was cleared one item at a time as the speech interface generated an appropriate sentence. Figure 5 illustrates the linkages that exist between all the components that allow this sequence to take place. The greyed elements show optional pathways supporting user interaction and query.



EARS: Old College is in front of you to your left at 110 m
EARS: Adam House is directly in front of you at 50 m

User: *Adam House*

EARS: OK – I heard Adam House
Some info – University of Edinburgh building used for examinations, matriculation, and careers shows.

User: *Old College*

EARS: OK – I heard Old College
Some info – Home to the Talbot Rice gallery, many of the University offices, and law course. The Playfield library in particular has spectacular décor. Notice the golden figure on top of the dome.

EARS: Royal Museum of Scotland is in front of you slightly to your left at 150 m

User: *Audiotags*

EARS: There are currently 3 audiotags available – Old College, Adam House, Royal Museum of Scotland

EARS: Sheriff court is directly in front of you at 160 m

User: *Closest point of interest*

EARS: The closest point of interest is the Royal Museum of Scotland at 80 m

EARS: George Herriot School is just visible directly in front of you at 560 m

User: *Add waypoint "Chambers Street"*

EARS: Waypoint added

EARS: Greyfriar's Bobby Statue is just visible in front of you at 100 m

User: *Greyfriar's Bobby Statue*

EARS: OK – I heard Greyfriar's Bobby Statue
Some info – The story goes that for 14 years a Skye T....

User: *Shut up*

EARS: OK – list cleared

Figure 6 Example of system output (3D CAD data copyright EWHT)
This figure appears in colour in the electronic version of this article and in the plate section at the back of the printed journal

Figure 6 is an example of a conversation between the user and the system, as the user travels down Chambers Street in Edinburgh. Various elements of the system are demonstrated, including recording of waypoints, and requests for further information on announced FoI.

8 Evaluation

The performance of *EARS* was assessed during 12 hours of field testing and five hours of user testing. During this time GPS and dialogue history was recorded to log files, for post

Table 5 Example of optional words and confidence level balancing in a recognition tag

POI ID	Audiotag	Recognition Tag
62	institute of geography	?institute ?of + geography

test analysis – an idea suggested by Dybkjær et al. (1995). In all 12,000 GPS signal strength recordings from around the city were captured. The system functions evaluated were:

- Effectiveness of speech based interfaces for delivery of spatial information;
- The accuracy of the viewshed data;
- GPS performance in urban environments;
- Usability/'wearability' of *EARS*.

8.1 Speech performance

The user tests included speakers of differing nationalities to assess the application's ability to cope with accents. The initial tests showed that it was capable of recognising commands from the active grammar without any user voice profile training, while in a quiet environment. However when background street noises were present the system performed much better for those users who had carried out five minutes of voice profile training. It was found that one hour of training produced markedly better performance in the noisiest of conditions. The interface had some flexibility to cater for human vagueness, which Norman (1998) terms as 'being analogue in a digital world'. This was achieved by setting up optional words in grammar rules, by using a "+", "–", and "?"(as can be specified in MS SAPI 5.1) to respectively increase or decrease confidence, or to make words optional. This adjustment can be carried out on individual words. In the example of Table 5, the user could say either 'Institute Geography', 'Geography', or 'Institute of Geography' to trigger a request for more information on this FoI.

This solved the problem in most cases, but unfortunately a tradeoff of increasing the confidence level was that greater training was required to maintain a high recognition accuracy. The system was confused by certain accents for the commands "enable" (activating the machine) and "disable" (placing the system in sleep mode). Therefore the commands were changed to the more differentiable "pay attention" and "disable", which improved recognition accuracy. The speech synthesizer (that spoke the text derived from the gazetteer) was considered to be understandable by all users.

8.2 Viewshed Performance

There was considerable flexibility in how relative importance of factors can be varied (in terms of announcement order, time interval between repetition, etc). Figure 7 was produced using the default settings of the system. To illustrate the order in which FoIs were announced, the FoIs of interest are numbered (numbers between 3 and 86). The user walked the route indicated by the arrows. The respective building was announced at the point indicated by the circled numbers (46, 75, 47 etc).

Two interesting observations are made: the system chose to announce FoI 75 before it announced 47 (a building that was closer but was just out of view at this time). Secondly, building 33 was not announced until the user turned east (on Edinburgh's

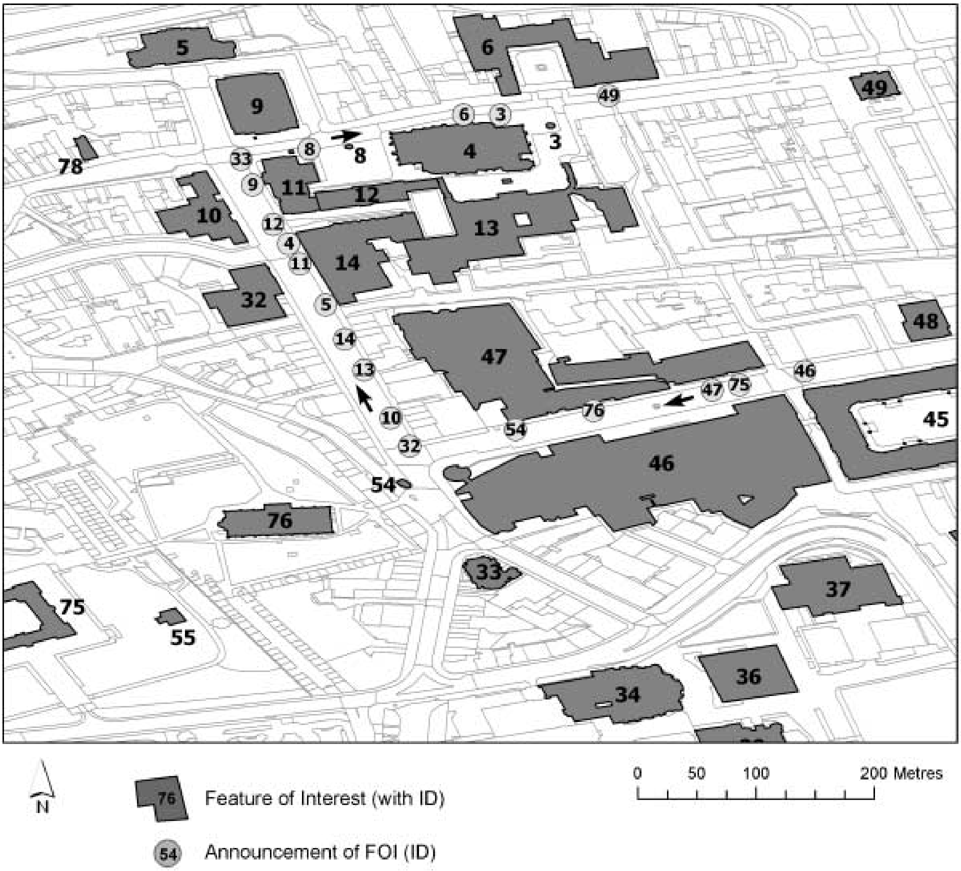


Figure 7 Map showing locations at which FoI (corresponding numbers) were announced (MasterMap data, Ordnance Survey © Crown copyright. All rights reserved)

Royal Mile), quite some distance from building 33. This was because the system was filtering out the announcement due to the FoI being outside of the ±90 degrees from direction of travel (i.e. behind user). However at the right turn onto the Royal Mile building 33 would have been within the range to be considered 'in front of', or to the 'side of the user', and still sufficiently close to be worthy of mention. Tests showed that on occasion a FoI would be announced but not be visible. Further analysis revealed that this was a result of vegetation or remaining residue from the building masking process. This is examined in more detail later in this paper.

8.3 GPS performance

GPS locational accuracy was a function of the satellite clock and orbital variation, atmospheric variations, receiver noise, and multipathing of signals (Kleusberg and Langley 1990). While many introduce a constant error in any given urban environment, the occlusion of satellite broadcasts and multipathing will vary across the city (Beesley 2003). The user's position on the street also has an effect on their field of view of the sky. A user walking on the pavement on a street with buildings 30 m high either side will typically have a skyward viewing angle of 45 degrees. When they turn the corner at a road intersection, the amount of visible sky could increase to 120 degrees. At this point typically more satellites become available and the signal strength increases.

The log files showed the most problematic areas tended to be where the user turned a corner. In some cases the GPS would continue to record coordinates for the previous direction of travel, akin to inertia in the system, taking a while to update to the new direction. This can be seen in Figure 8 where the user turned left walking past Point A, but the GPS continued the track to Point B before correcting itself. The reason for this is illustrated in Figure 9 which shows a change in satellite availability (total number, and satellite ID) as the user turns a corner.

The variation in signal strength arising from the effect of satellite occlusion can be seen in Figure 10 as the user crosses a raised bridge. The symbol size is proportional to the signal to noise ratio, larger symbols indicating a more powerful GPS signal. A further test using 122 samples showed that in the urban context the GPS device was able to offer an average of 12 m horizontal precision, while in the open space (such as Holyrood park in Edinburgh City), a precision of 5.7 m was possible. Signal degradation was evident under scaffolding, bus shelters, and vegetation canopies. GPS tracks were plotted and overlaid with LIDAR data to assess the source of the error, and the degree of impact. Dense trees sometimes led to complete loss of signal (the track 'jumping' a distance of 10 m or so), whilst in other situations, there was deterioration in the precision of the location given (typically from a precision of 9 to 14 m). A concern was that this locational error could result in inappropriate announcements. For example in Figure 11, the GPS reported the position of user 'A' as being at 'B' – thereby announcing FoIs not actually visible to the user.

Alternatively the GPS might give the correct position, but the LIDAR datasets place the user higher than they were in reality. This arose where the user passed under vegetation canopy. Tree height was captured in the LIDAR and when calculating the viewshed, the user, was in effect, placed on top of the canopy. Again, FoIs were reported as being visible when in reality they were not. For this reason the system was modified early in the testing phase such that it compared the LIDAR elevation data for the current GPS latitude and longitude, with that reported by an altimeter fitted inside the GPS. Before each walk the altimeter would be set to a known benchmark height taken from the LIDAR dataset. At any time during a walk if the discrepancy was greater than 16 m the search of viewshed data was halted. Should ten searches in a row be halted then the system would notify the user. This discrepancy threshold of 16 m was determined through analysis of log files of announcement and comparison with height readings from the GPS and LIDAR estimated heights along various tracks.

8.4 Wearability of the System

For all users the system was the first experience they had had of a mobile augmented reality system. Comments were very favourable, one commenting 'I think speech-based systems offer better interfaces for mobile devices than visual interfaces'. The system ran on a laptop inside a small back pack, with an ear piece and small microphone arm. Even this was found to be 'a little obtrusive', prompted another to say: 'It would be great if it could be smaller and lighter'. At some points along a journey, background noise was sufficient to drown out *EARS*, one user suggesting 'it sometimes speaks too quickly, and

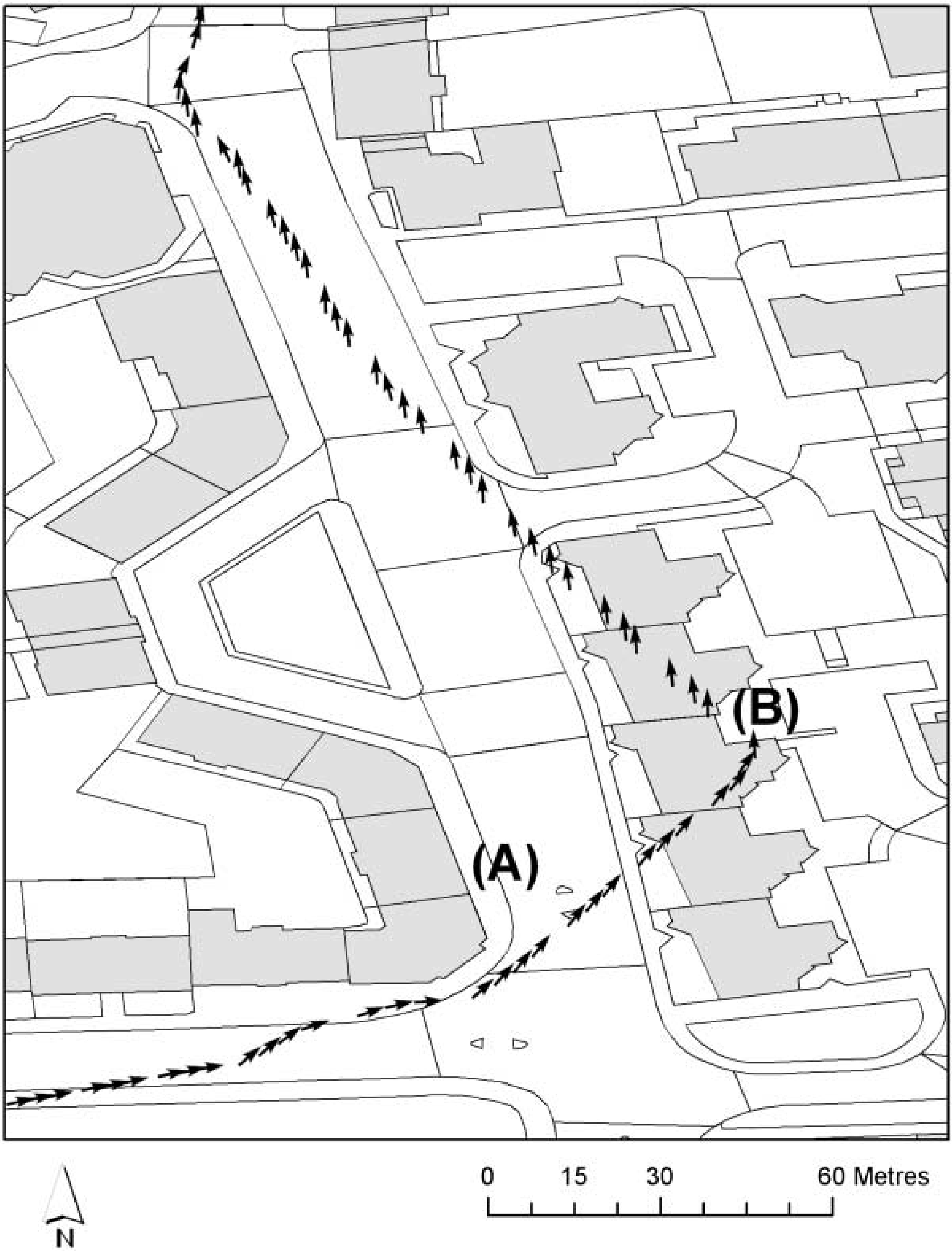


Figure 8 GPS tracking issues on corners (MasterMap data, Ordnance Survey © Crown copyright. All rights reserved)

it would be useful to ask it to repeat the last item in case you didn't hear it . . . like when a bus was passing by'. It was noted that speech accuracy could be improved if the user stood still while making the announcement, as this reduced the wind noise in the microphone. It was also suggested that the system could provide a confirmatory role by

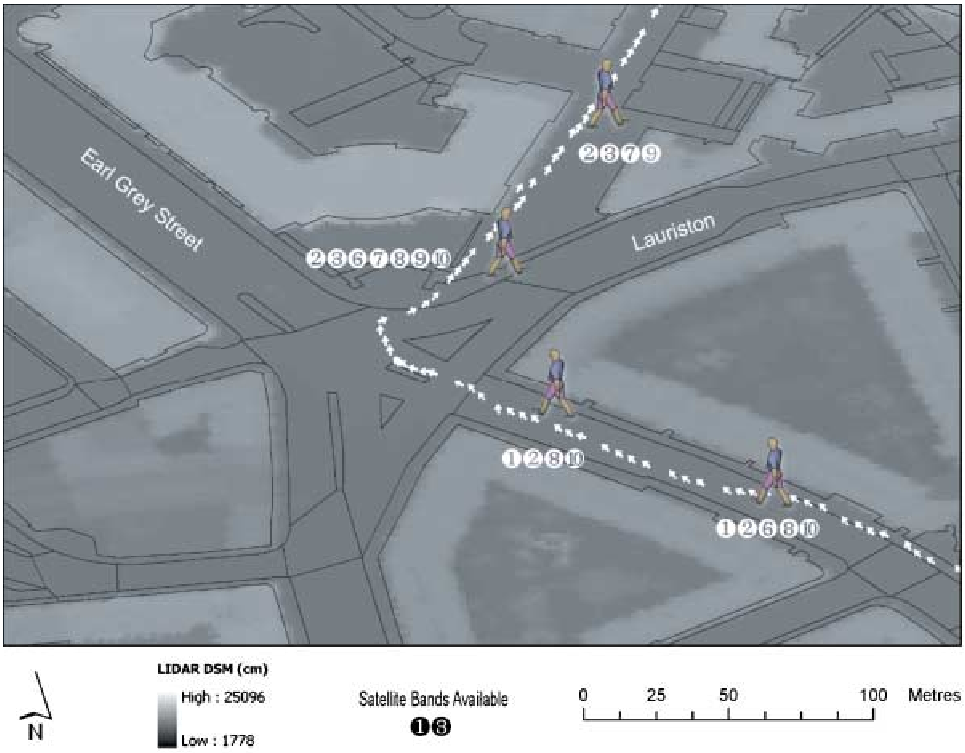


Figure 9 GPS satellite availability (LIDAR data copyright Environment Agency, MasterMap data, Ordnance Survey © Crown copyright. All rights reserved)
This figure appears in colour in the electronic version of this article

re-announcing the FoI when the user was near to it, saying that the user was "now at" the FoI. Since it was not always clear what the user should be able to see, one user suggested that a description of the FoI should be included so that it may be identified more easily in busy urban horizons (e.g. "Old College has a dome with a golden figure on top"), or that description of adjacent buildings might be used to reference the user.

It was also noted that description in terms of the relative angle and approach to the FoI was sometimes dependent on where and how coordinates were used to record the feature of interest. Since each FoI was represented as a point feature, there are times when an announcement can incorrectly state that the FoI was in front/behind the user while it was actually alongside them (Figure 12). To overcome this *EARS* should use polygonal data to represent each FoI, and report distances and angles based on the closest edge. When a user was close to the FoI the phraseology would reflect this, reporting that the building was 'to your right' as the user walked alongside the FoI.

Future developments might include the use of a digital compass to provide orientation, though our research revealed digital compasses to be insufficiently reliable because they are too sensitive to variations in their inclination with respect to magnetic north.

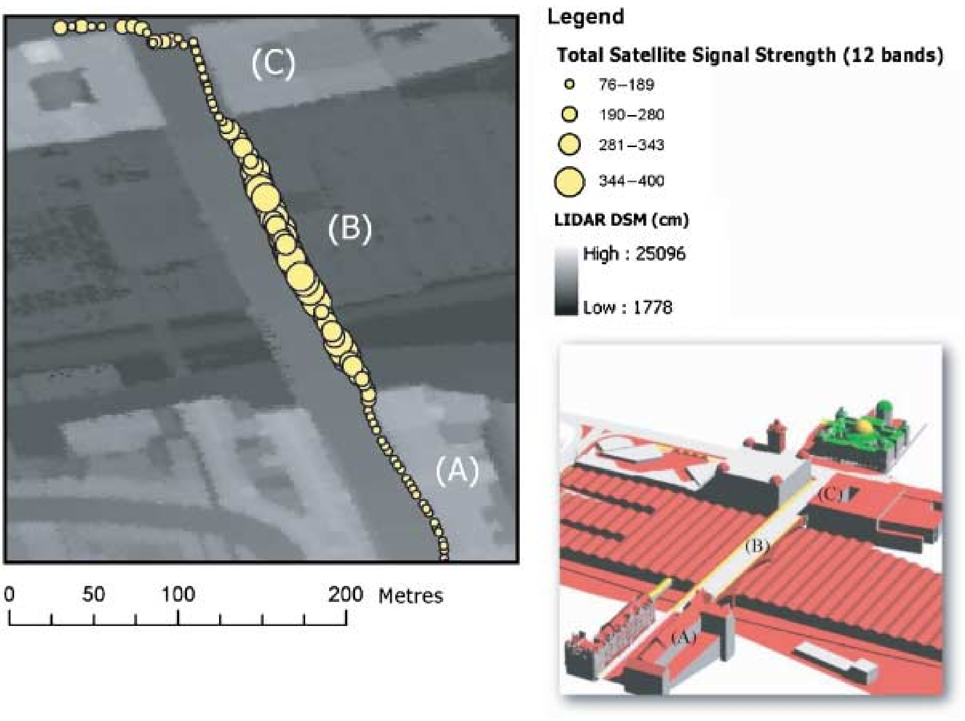


Figure 10 GPS signal strength (LIDAR data copyright Environment Agency, 3D CAD data copyright EWHT)
This figure appears in colour in the electronic version of this article and in the plate section at the back of the printed journal

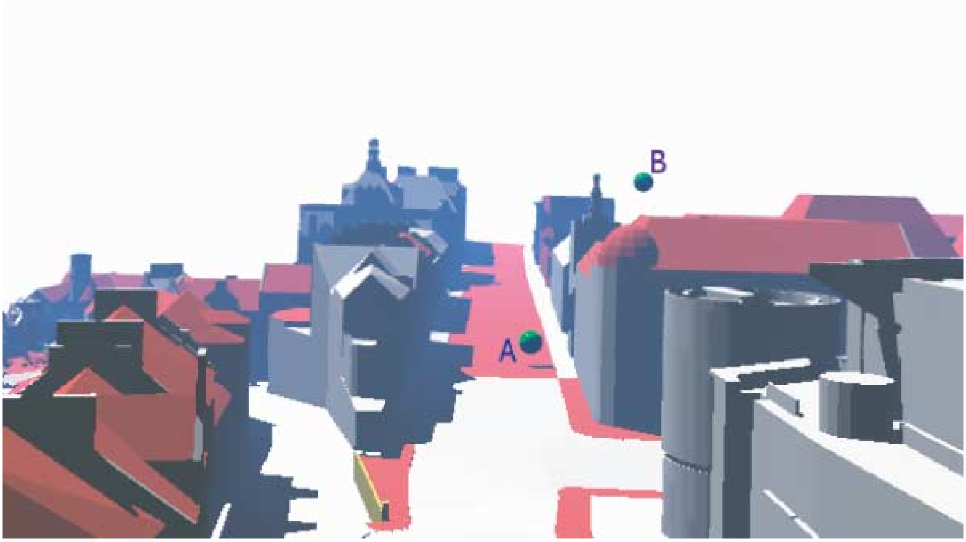


Figure 11 Significance of minor horizontal positioning errors (3D CAD data copyright EWHT)
This figure appears in colour in the electronic version of this article and in the plate section at the back of the printed journal

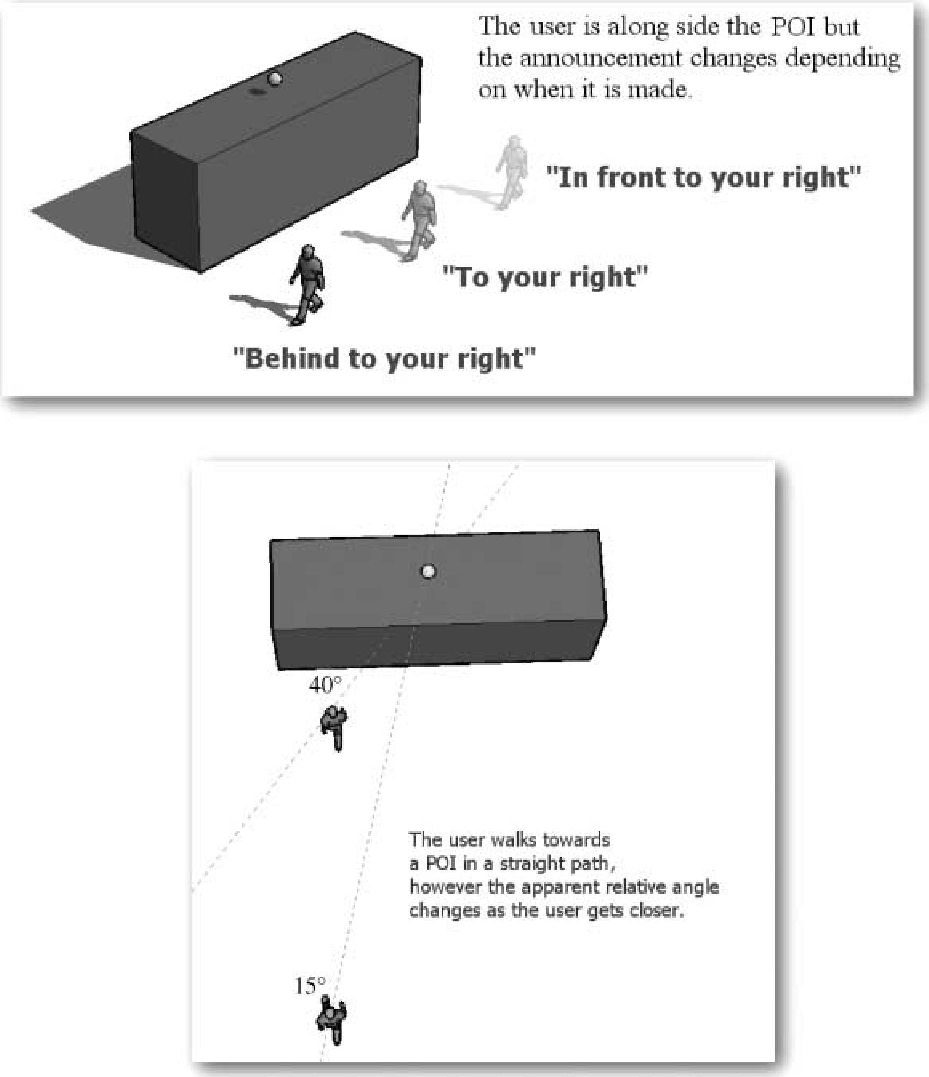

The user is along side the POI but
the announcement changes depending
on when it is made.
"In front to your right"
"To your right"
"Behind to your right"
40°
The user walks towards
a POI in a straight path,
however the apparent relative angle
changes as the user gets closer.
15°

Figure 12 Issues regarding the use of a centroid to represent a 3D object

9 Conclusions

This research has realized its aims and objectives by successfully implementing a system which delivers, via a speech-based interface, contextual geospatial information in real time to assist a user in locating landmarks within an urban environment – in essence delivering qualitative descriptions of space using quantitative techniques. The system differs from current LBSs in two ways. First, it delivers information based on the degree of visibility of an item, not just its proximity, from the user's location. It was intentional that *EARS* did not provide any wayfinding functionality, but instead simply announced key landmarks as they become visible, allowing free exploration as a way of discovering the city (Edwardes and Mackaness 2001). Secondly, whereas existing systems tend to augment information visually, this project focused on delivery and interaction with information via the spoken word. In the context of ubiquitous computing and location-based services, the challenge has been in developing speech-based interactions between the user and the machine that move beyond a series of discrete steps and prompts to one of a continuous conversation – one that instantaneously responds, in a non intrusive, non prescriptive manner. *EARS* was a 'dynamically context-aware' (Stephanidis 2003) application, and allowed the user a hands free, as well as 'eyes free' (augmented information does not interfere with the user's field of view) in order that the user can explore and augment their knowledge of their surroundings in a natural way.

It has also demonstrated that standard GIS datasets are suitable for the construction of the necessary visibility databases for such a system, and that WAAS enabled GIS provides sufficient accuracy for user location in the urban environment, though improved location solutions should be available in the near future. Spatial indexing has addressed the important issue of speed in information retrieval from large 2.5D datasets. The value of field evaluation was highlighted during the design life cycle of the system, as was the suitability of LIDAR for calculating the visibility of target points in an urban environment.

The target audience for *EARS* was tourists visiting an area for a limited period of time who wished to use a virtual city guide to point out interesting features. It could be used by locals wishing to learn more about their surroundings. Beyond just tourism, there are other avenues for this type of technology:

- Virtual walk throughs – planning a visit prior to arriving at a city, or exploring virtual worlds;
- Exploring rural environments – description of rural landscapes in normal/poor visibility;
- Visually Impaired – warning of recent obstructions or navigation points;
- Spatial Note Taking – Police crime scene note taking, archaeology site notes;
- Military applications – voiced information on content and use of strategic buildings; and
- Gaming – mixed virtual and augmented reality among multiple participants.

Acknowledgments

The authors are very grateful to The Environment Agency for the supply of the LIDAR datasets, to the Ordnance Survey for sections of *MasterMap* data available via the CHEST agreement, to Mr Jon Mengham of the Edinburgh World Heritage Trust for use of their 3D CAD dataset of the Old Town of Edinburgh, to Mr Bruce Gittings for permitting the use of sections of the *Gazetteer for Scotland*, and to reviewers for their suggested improvements.

References

Bartie P J 2004 Development of a Speech Based Augmented Reality System to Support Exploration of Cityscape (Edinburgh as a Case Study). Unpublished MSc Thesis, Institute of Geography, University of Edinburgh

Beesley B J 2003 Sky Viewshed Modelling for GPS Use in the Urban Environment. Unpublished M.S. Thesis, Department of Geography, University of South Carolina

Dale R, Geldof S, and Prost J-P 2003 CORAL: Using natural language generation for navigational assistance. In Oudshoorn M (ed) *Proceedings of the Twenty-sixth Australasian Computer Science Conference, Adelaide, Australia*. Sydney, Australian Computer Society: 35–44

Denis M, Pazzaglia F, Cornoldi C, and Bertolo L 1999 Spatial discourse and navigation: An analysis of route directions in the City of Venice. *Applied Cognitive Psychology* 13: 145–74

Dybkjær L, Bernsen N O, and Dybkjær H 1995 Different spoken language dialogues for different tasks: A task-oriented dialogue theory. In Pfleger S, Gonçalves J, and Varghese K C (eds) *Advances in Human-Computer Interaction: Human Comfort and Security*. Berlin, Springer: 46–61

Edwardes A and Mackaness W A 2001 Mobile mapping on demand: Redefining the needs of cartographic generalisation. In *Proceedings of the Ninth Annual GISRUK Conference*, Pontypridd, Wales: 333–5

Egenhofer M and Shariff A R 1998 Metric details for natural-language spatial relations. *ACM Transations on Information Systems* 16: 295–321

Feiner S, Höllerer T, Gagas E, Hallaway D, Terauchi T, Güven S, and MacIntyre B 2004 MARS: Mobile Augmented Reality Systems. WWW document, http://www1.cs.columbia.edu/graphics/projects/mars/mars.html

Feiner S 2002 A New Way of Seeing. WWW document, http://www.sciam.com/article.cfm?articleID=0006378C-CDE1–1CC6-B4A8809EC588EEDF&pageNumber=1&catID=2

Frank A 1992 Qualitative spatial reasoning about distance and directions in geographic space. *Journal of Visual Language and Computing* 3: 343–71

Hirtle S C and Heidorn P B 1993 The structure of cognitive maps: Representations and processes. In Gärling T and Golledge R G (eds) *Behavior and Environment: Psychological and Geographical Approaches*. Amsterdam, North Holland: 1–29

Kleusberg A and Langley R B 1990 The limitations of GPS. *GPS World* 1(2): 50–2

Levinson S C 1996 Frames of reference and Molyneux's Question: Crosslinguistic evidence. In Bloom P, Peterson M A, Nadel L, and Garrett M F (eds) *Language and Space*. Camridge, MA, MIT Press: 109–69

Lovelace K L, Hegarty M, and Montello D R 1999 Elements of good route directions in familiar and unfamiliar environments. In Freksa C and Mark D M (eds) *Spatial Information Theory*. Berlin, Springer Lecture Notes in Computer Science No. 1661: 65–82

Lynch K 1960 *The Image of the City*. Cambridge, MA, MIT Press

Maybury M 1993 *Intelligent Multimedia Interfaces*. Cambridge, MA., AAAI Press/MIT Press

Michon P-E and Denis M 2001 When and why are visual landmarks used in giving directions? In Montello D R (ed) *Spatial Information Theory*. Berlin, Springer Lecture Notes in Computer Science No. 2205: 292–305

Microsoft 2003 *Microsoft SDK – Speech Application Programmers Interface Version 5.1 Help File*. Redmond, WA, Microsoft Corp.

Mineter M, Dowers S, Gittings B, and Caldwell D 2004 A Multi-computing Software Environment for ArcInfo Visibility Analysis. WWW document, http://gis.esri.com/library/userconf/proc02/pap1253/p1253.htm

mySQL 2004 Server Performance Tuning (Chapter 6). WWW document, http://dev.mysql.com/books/hpmysql-excerpts/ch06.html

Norman D 1998 *The Invisible Computer*. Cambridge, MA, MIT Press

Palmer T C and Shan J A 2002 Comparative study on urban visualization using LIDAR data in GIS. *URISA Journal* 14: 19–25

Pashler H 1995 Attention and visual perception: Analyzing divided attention. *Visual Cognition* 2: 1–70

Pilgrim M 2002 What was RSS? WWW document, http://www.xml.com/pub/a/2002/12/18/dive-into-xml.html

Rauterberg M 2002 History of HCI. WWW document, http://www.ipo.tue.nl/homepages/mrauterb/presentations/HCI-history/sld096.htm

Reiter E and Dale R 2000 *Building Natural Language Generation Systems*. Cambridge, Cambridge University Press

Revell S 2001 How does Location Govern task? Unpublished MSc Thesis, Institute of Geography, University of Edinburgh

Rottensteiner F and Briese C 2002 A new method for building extraction in urban areas from high-resolution LIDAR data. *International Archives of Photogrammetry and Remote Sensing* 34: 295–301

Salmon R and Slater M 1987 *Computer Graphics: Systems and Concepts*. Wokingham, Addison-Wesley

Stephanidis C 2003 Towards universal access in the disappearing computer environment. *Upgrade* 4: 53–9

Tobler W 1970 A computer movie simulating urban growth in the Detroit region. *Economic Geography* 46: 234–40

Tom A and Denis M 2003 Referring to landmark or street information in route directions: What difference does it make? In Kuhn W, Worboys M, and Timpf S (eds) *Spatial Information Theory*. Berlin, Springer Lecture Notes in Computer Science No. 2825: 384–97

Tomlin D 1990 *Geographic Information Systems and Cartographic Modelling*. Englewood Cliffs, NJ, Prentice-Hall

Tversky B 1993 Cognitive maps, cognitive collages, and spatial mental models. In Frank A U and Campari I (eds) *Spatial Information Theory*. Heidelberg, Springer Lecture Notes in Computer Science No. 716: 14–24

Weber C R 1998 The representation of spatio-temporal variation in GIS and cartographic displays: The case for sonification and auditory data representation. In Egenhofer M and Golledge R (eds) *Spatial and Temporal Reasoning in Geographic Information Systems*. New York, Oxford University Press: 74–84