ORIGINAL ARTICLE

NAVIG: augmented reality guidance system for the visually impaired

Combining object localization, GNSS, and spatial audio

**Brian F. G. Katz · Slim Kammoun · Gaëtan Parseihian · Olivier Gutierrez ·
Adrien Brilhault · Malika Auvray · Philippe Truillet · Michel Denis ·
Simon Thorpe · Christophe Jouffrais**

Abstract Navigating complex routes and finding objects of interest are challenging tasks for the visually impaired. The project NAVIG (Navigation Assisted by artificial VIsion and GNSS) is directed toward increasing personal autonomy via a virtual augmented reality system. The system integrates an adapted geographic information system with different classes of objects useful for improving route selection and guidance. The database also includes models of important geolocated objects that may be detected by real-time embedded vision algorithms. Object localization (relative to the user) may serve both global positioning and sensorimotor actions such as heading, grasping, or piloting. The user is guided to his desired destination through spatialized semantic audio rendering, always maintained in the head-centered reference frame. This paper presents the overall project design and architecture of the NAVIG system. In addition, details of a new type of detection and localization device are presented. This approach combines a bio-inspired vision system that can recognize and locate objects very quickly and a 3D sound rendering system that is able to perceptually position a sound at the location of the recognized object. This system was developed in relation to guidance directives developed through participative design with potential users and educators for the visually impaired.

Keywords Assisted navigation · Guidance · Spatial audio · Visually impaired assistive device · Need analysis

B. F. G. Katz (✉) · G. Parseihian · M. Auvray · M. Denis
LIMSI-CNRS, Université Paris Sud, 91403 Orsay, France
e-mail: brian.katz@limsi.fr

G. Parseihian
e-mail: gaetan.parseihian@limsi.fr

M. Auvray
e-mail: malika.auvray@limsi.fr

M. Denis
e-mail: michel.denis@limsi.fr

S. Kammoun · O. Gutierrez · A. Brilhault · P. Truillet ·
C. Jouffrais
IRIT, CNRS & Université Paul Sabatier, Toulouse, France
e-mail: kammoun@irit.fr

O. Gutierrez
e-mail: gutierrez@irit.fr

A. Brilhault
e-mail: adrien.brilhault@cerco.ups-tls.fr

P. Truillet
e-mail: truillet@irit.fr

C. Jouffrais
e-mail: jouffrais@irit.fr

A. Brilhault · S. Thorpe
CerCo, CNRS & Université Paul Sabatier, Toulouse, France
e-mail: simon.thorpe@cerco.ups-tls.fr

1 Introduction

According to many reports issued by national committees [see e.g. Ińegalíte des Chances (Canadian Institute for the Blind 2005)], visually impaired—including blind and low vision—people require assistance in their daily life. One of the most problematic tasks is navigation, which involves two main action components—mobility and orientation. These two processes were defined by Loomis et al. (2006). The first element relates to sensing the immediate (or near-field) environment, including obstacles and potential paths in the vicinity, for the purpose of moving through it. It may rely on visual, auditory, or olfactory stimuli identification and localization. In the literature, this set of processes has

been labeled as Micro-Navigation or Mobility. The second component, typically termed Orientation or Macro-Navigation, includes multiple processes such as being oriented, selecting an appropriate path, maintaining a path, and detecting when the destination has been reached. All these tasks are dedicated to processing the remote (or far-field) environment, beyond the immediately perceptible one. In the case of visual impairment, the main cues used for sensing the environment (e.g., detecting obstacles, landmarks, and paths) are lacking. As such, both micro- and macro-navigation processes are degraded, resulting in difficulties related to obstacle avoidance, finding immediate paths, piloting (guidance from place to place using landmarks), correct orientation or heading, maintaining the path, etc.

A recent literature review of existing electronic aids (2007–2008) for visually impaired (VI) individuals identified more than 140 products, systems, and assistive devices while providing details on 21 commercially available systems (Roentgen et al. 2008). The different systems were divided into two categories: (1) obstacle detection or micro-navigation and (2) macro-navigation. The different cited micro-navigation systems are mainly devoted to improving obstacle avoidance (by extending the range of the white cane) or "shorelining". They do not provide assistance for object localization or more complex navigation tasks.

Macro-navigation systems are almost exclusively based on the use of Global Positioning System (GPS) with adaptations for visually impaired users (see Loomis et al. 2005; Marston et al. 2006 for an evaluation overview of GPS devices for VI individuals). The MoBIC (Strothotte et al. 1995) project (mobility for blind and elderly people interacting with computers), presented in 1995, addressed outdoor navigation problems for the visually impaired. The MoBIC system consisted of two components: the MoPS pre-journey system that enables preparation at home, and the MoODS outdoor system that provides positioning and easy access to all necessary information. Evaluations have shown that such a system is primarily limited by the precision of the positioning system and the details in the geographic database. Another very important project in the area of navigation without sight was conducted over a period of twenty years by Loomis, Golledge, and Klatzky. Their goal was the development of a personal guidance system (PGS) for the blind (Loomis et al. 1994). The PGS system included three modules: (1) a module for determining the position and orientation of the traveler using Differential GPS (DGPS), (2) a Geographic Information System (GIS) module, and (3) a user interface. Several evaluations were performed, which demonstrated the effectiveness of such a device in helping visually impaired persons navigate.

Other works include (Ran et al. 2004), which extended the outdoor version of Drishti (Helal et al. 2001), an integrated navigation system for the visually impaired and disabled, to a navigation assistive device integrating indoor positioning. Indoor positioning was accomplished using an ultrasound device. Results showed an indoor accuracy of 22 cm. Concerning outdoor positioning, DGPS was recommended. Differential GPS improves GPS accuracy to approximately ± 10 cm in the best conditions. Unfortunately, it relies on a network of fixed, ground-based reference stations that are currently only available in North America (via the Wide Area Augmentation System).

Outside of research related projects, several adapted GPS-based navigation systems for VI users have been commercialized, with probably the most popular GPS-based personal guidance systems being **BrailleNotesGPS** (Sendero Inc.) and **Trekker** (HumanWare Inc.). Although they are very useful in unknown environments, they still suffer usability limitations (especially due to GPS precision and map weakness). While no commercial products that were able to detect and locate specific objects without the necessity of pre-equipping them with dedicated sensors (e.g., RFID tags) were reported, several research projects are considering this issue. A visual-based system, incorporating a handheld stereo camera and WiFi-based tracking for indoor use was presented in Hub et al. (2004). This system relies on the use of 3D models of precise objects and predefined spaces in order for them to be identified, greatly limiting its use outside of the designed environment. Direct sensory substitution systems, which directly transform data from images to auditory or tactile devices without any interpretation, can be used for rudimentary obstacle detection and avoidance, or for research on brain plasticity and learning (Auvray and Myin 2009; Auvray et al. 2007). However, these systems have a large learning curve and are very hard to use in any practical sense.

Thus, it appears that micro-navigation devices focus on obstacle detection only and that macro-navigation devices are mainly regular GPS systems with the addition of non-visual interactions. None of the assistive devices mentioned have aimed to improve the sensing of the immediate environment, which is a fundamental process in navigation and spatial cognition in general.

Based on the needs of VI individuals (Gallay et al. 2012; Golledge et al. 2004), the NAVIG[1] project (2009–2012) aims to design an assistive device that provides aid in two problematic situations: (1) near-field sensing (specific objects identification and localization) and guidance (heading, grasping, and piloting); and (2) far-field navigation relying on an adapted GIS database and route selection process adapted to VI individuals). For both

---

[1] Projet NAVIG, http://navig.irit.fr.

near-field and far-field tasks, guidance will be provided through the generation of an audio augmented reality environment via binaural rendering, providing both spatialized text-to-speech and sonifications, allowing the full exploitation of the human perceptual and cognitive capacity for spatial hearing (Dramas et al. 2008). The combination of these two functions provides a powerful assistive device. The system should permit VI individuals to move at their own pace toward a desired destination in a sure and precise manner, without interfering with normal behavior or mobility. Through the use of an artificial vision module, this system also assists users in the localization and grasping of objects in the immediate environment without the need to pre-equip them with electronic tags.

The aim of this paper is to present the NAVIG system architecture and the specifics related to macro and micro-navigational use. NAVIG follows a method of participatory design and some aspects of this process are presented here. A presentation of the first prototype is also shown, with a discussion of developments underway. Portions of this work have been previously presented (see Katz et al. 2010, 2012; Parseihian et al. 2010; Kammoun et al. 2010, 2012; Brilhault et al. 2011; Parlouar et al. 2009).

2 User needs and participatory design

The first point that must be emphasized is that the NAVIG system is not intended to replace the white cane or the guide dog. It should rather be considered as a complement to these devices and general Orientation and Mobility (O&M) training sessions. Therefore, the primary objective for the device is to help VI people (i.e., early and late blind as well as people partially deprived of vision) improve their daily autonomy and ability to mentally represent their environment beyond what is possible with traditional assistive devices. The visually impaired most often employ egocentric spatial representation strategies (Gallay et al. 2012; Noordzij et al. 2006; Afonso et al. 2010). However, the integration of several different paths into a general or global representation is necessary for route variations, such as detours, shortcuts, and journey reorganization. Creating these internal representations requires additional effort on the part of the individual. A study of the cognitive load associated with the use of language and virtual sounds for navigation guidance can be found in (Klatzky et al. 2006).

Designing an assistive device for VI users implies the need to sufficiently describe the problems these individuals typically face and hence the needs that the device should respond to. In order to provide ongoing assistance throughout the project, the NAVIG team included the Institute for the Young Blind, Toulouse, enabling a participatory design strategy with potential users of the

system. A panel of four O&M instructors and a number of VI participants were involved in brainstorming, participatory design sessions, as well as psychological and ergonomic experiments and prototype evaluations. The panel comprised 21 VI potential users (6 female/15 male, aged 17–57, mean 37). A questionnaire was given concerning their degree of autonomy and technological knowledge. For daily mobility, five of the panel employed a guide dog, 12 used a white cane, while the remaining 4 preferred to have a person to guide them. A product of the various design sessions with O&M instructors and visually impaired volunteers has been the construction of a number of guidelines for the development of the assistive device, whose main results are provided below (see Brunet 2010 for further details).

2.1 Route planning study

A brainstorming session with six VI panel members was conducted to address the needs for journey planning for an autonomous pedestrian itinerary (Brunet 2010). As the study aimed at generalizing findings to both early and late blind people, a mixture of participants was included. Results indicated that a detailed preliminary planning phase improved cognitive representations of the environment.

The need for a customizable preliminary planning phase was then confirmed by an empirical study focusing on the interconnection between the preparation and execution phases of navigation and on the internal and external factors underlying this activity. Six visually impaired (4 early blind and 2 late blind, ages 27–57, mean 38.9) Parisian participants were followed by experimenters while preparing and completing a 2-km unfamiliar route in a residential area of Paris, France, (see Fig. 1) using the technique of information on-demand (adapted from Bisseret et al. 1999). This experiment consisted of allowing the participants to ask the experimenter the desired information instead of searching for it by
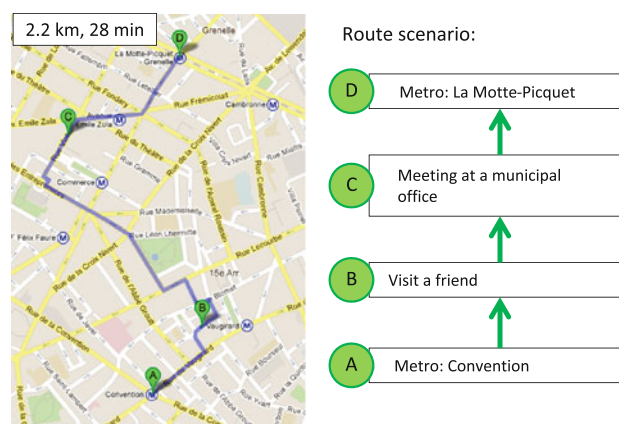

Fig. 1 Scenario used for route planning study

Table 1 Group descriptions and group means of the number of questions posed during the *information on-demand* experiment during the preparatory (Prep.) and route exploration (Route) phases

Group	Preparation condition	Number of questions		
		Prep.	Route	Total
G1	Without human assistance	40	146	186
G2	With human assistance	77	156	233
G3	No preparation	0	207	207

themselves, so that the experimenter simulated an ideal navigation assistive device.

Participants were divided into three groups according to the level of preparatory planning performed (preparation with human assistance, without human assistance, and no preparation at all). A description of the different groups and the number of questions posed by each group to the experimenter is given in Table 1 for both the preparatory and route guidance phases. Analysis of the transcribed questions allowed for the classification of the different questions posed by all participants. Several categories of questions could be defined, allowing them to be classified according to their relationship to the state of progress along the route. The occurrence of the different categories of questions with respect to the different groups highlights the relation between route planning and the type, and the amount of required information for an efficient and pleasant trip, in addition to the category descriptions, is shown in Table 2.

The amount of information requested by the participants to complete the journey, combined with the analysis of post-experimental interviews, was used to define several user profiles. Each profile consisted of different strategies

Table 2 Classification of questions posed during the route guidance phase of the *information on-demand* experiment according to trajectory segment, **T**, relative to the current point **T0**, and the proportion of each question category as a function of experimental group **G**

Category	Question pertaining to	G1 (%)	G2 (%)	G3 (%)	Overall (%)
T-1	Previous route segment	3	3	3	3
T0	A precise point	34	24	27	28
T1	Current route segment, next step	48	32	38	38
T2	The next 2 route segments	11	9	12	10
T3-4	The next 3 to 4 route segments	4	11	14	10
T5	More than 4 segments in advance, global route planning (global route concept, not precise instructions)	0	21	6	11

and as a result had a specific set of needs that should be separately addressed by the NAVIG system.

This design study thus allows for the creation of guidelines related to the guidance information that is to be provided to the user during the course of navigation, taking into account the presence, or absence, of preparatory route planning. In particular, the advantages observed for users who performed preparatory route planning have led to the definition of an overview function for the guidance system. Users of the NAVIG system should be able to prepare and preview their upcoming journey at home, to access an overview or a more detailed description of the path to follow, to receive knowledge about the different points of interest along the path including appropriate spatial cues to locate their specific path with respect to the general surrounding environment, and to be able to preview the actual route guidance.

Information relative to a route segment, and when to present this information, is a component of the route guidance system. A certain degree of flexibility in these rules should be designed into the system. The adjustment of these parameters will be the subject of further testing with the actual system.

2.2 Route difficulty assessment

Instructors and potential users have specifically insisted on the fact that an assistance device must be highly customizable. Concerning the routes offered by the system (see Sect. 3.6), this implies that the most independent and confident users can request the shortest path, even if it involves various difficulties, whereas the most conservative users can request a longer but easier to follow more prudent path route.

A means of identifying or classifying route points to favor or avoid when calculating an optimal route was seen as necessary. A 2-h brainstorming session was conducted with six panel participants. A consensus set of preferable route elements was found: simple crossings, roads with lighter traffic, larger walkways to allow for faster walking, and the shortest route if time is a chosen consideration. In addition, a set of elements to avoid was also established: plazas, roundabouts, large open areas, very large walkways (due to the increased presence of obstacles), very narrow walkways, shared pedestrian/automobile areas, and areas with many poles or fences. These findings converge with those by Gaunet and Briffault (2005).

In the interest of refining these results, in order to be able to apply them to automated route selection, a method to associate difficulty scores for the most common urban elements was established. Participants were first asked to cite the three types of events/obstacles they find the most problematic in a pedestrian journey. A list of items was derived from their answers. Participants were then asked to

rate the difficulty of each element on a scale from 1 to 5 (see Table 3). Those scores can be used to create a list of proposed paths which users will be able to select from, based on their own criteria of confidence, time available, and acceptable level of difficulty.

2.3 Ideal guidance information

Analysis of discussions concerning the audio guidance provided by an ideal system highlighted the need to control the amount of information presented. The quantity of information should take into account current conditions as well as individual user needs and preferences. In general, the amount of information provided should be minimal, avoiding excess, presenting only what is necessary, and sufficient to aid the user (see also Allen 2000; Denis 1997). The information provided should be highly efficient and minimally intrusive. These guidelines have been taken into consideration in the modification of georeferenced information databases (see Sect. 3.4.1) and in the design of the audio feedback to the user (see Sects. 3.5 and 3.6).

Table 3 Predominant pedestrian obstacles encountered with associated mean difficulty scores with standard deviation (rated on a scale from 1 to 5)

Obstacle	Difficulty score
Narrow sidewalk	4.67 ± 0.8
Large spaces, parks	4.33 ± 0.5
High obstacle (above cane detection)	3.33 ± 1.4
Roundabout	3.17 ± 0.4
Intersection without traffic lights	2.83 ± 0.4
Wide sidewalk	2.67 ± 0.5
Sidewalk terrace/street café	2.17 ± 0.4
Stairway	2.00 ± 1.1

3 System design

The different objectives of NAVIG will be attained by combining input data furnished through satellite-based geolocalization and an ultra-rapid image recognition system. Guidance will be provided using spatialized audio rendering with both text-to-speech and specifically designed semantic sonification metaphors.

The system prototype architecture is divided into several functional elements structured around a multi-agent framework using a communication protocol based on the IVY middleware (Buisson et al. 2002). With this architecture, agents are able to dynamically connect to, or disconnect from, different data streams on the IVY bus. The general architecture of the system is shown in Fig. 2. The main operating elements of the NAVIG system can be divided into three groups: data input, user communication, and internal system control. The data input elements consist of a satellite-based geopositioning system, acceleration and orientation sensors, Geographic Information System (GIS) map databases, an ultra-rapid module processing images from head-mounted stereoscopic cameras, and a data fusion module. User communications are handled predominantly through a voice recognition system for input and an audio rendering engine using text-to-speech and conceptual trajectory sonification for output. In the following sections, the different system components will be presented and discussed individually.

3.1 The user interface

The user interface (UI) acts as the directing component between user's requirements and system functions. In
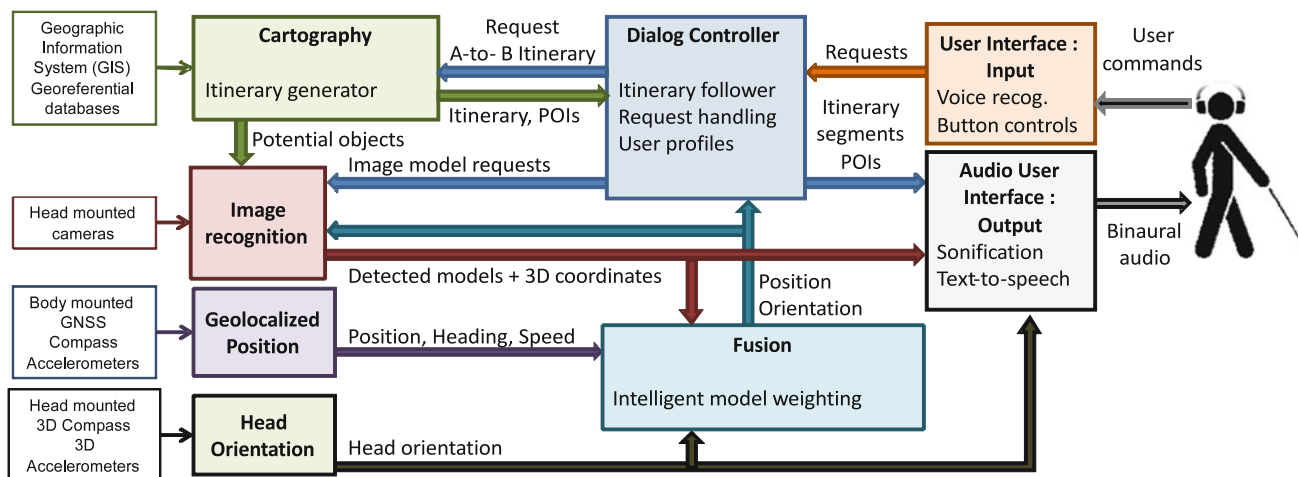

Fig. 2 NAVIG system architecture overview

Table 4 A prototypic address input scenario

User >	"Initiate navigation mode"
NAVIG >	"Outdoor navigation mode selected, please select your destination address, select an avenue"
User >	"Avenue Lazare Carnot"
NAVIG >	"You selected Avenue Lazare Carnot, select the avenue's number"
User >	"85"
NAVIG >	"Your destination is 85 Avenue Lazare Carnot"
User >	"Start navigation"

order to design an adapted UI, two separate brainstorming sessions were organized with VI individuals and system designers to develop a working scenario and interaction techniques. The goal was to define the nature and quantity of information required during a guided travel, as well as appropriate modalities that would not interfere with learned O&M techniques and abilities so as not be obtrusive. At the conclusion of these meetings, all participants cited speech interaction as the most natural and preferred method. An interactive menu was implemented based on a voice recognition engine (Dragon Naturally Speaking) and a text-to-speech generator (Elan Real Speak). The voice menu offers different possibilities for the user to interact with the NAVIG system in indoor and outdoor situations.

In both indoor and outdoor navigation, the user may ask for a specific object known to the system or for a destination to reach. Indoors, known objects may include stationary localized targets such as signs, elevators, or vending machines, as well as standardized objects such as furniture. If the map of the building is embedded, the destination may be a specific location or room. In outdoor situations, the main function consists in entering an address or place (e.g., postal office), as well as a known object (e.g., the mailbox, door entrance). An explicit strategy for dialog was implemented that allows a novice user to understand the menu hierarchy. Table 4 illustrates a typical address input, with the departure location provided by the actual geolocation of the user.

3.2 Object Identification and Localization

Improvements in computer hardware, the cheap cost of cameras, and increasingly sophisticated computer vision algorithms have contributed to the development of systems where the recognition and localization of visual features can be used in navigation tasks. The typical approach is to track a set of arbitrary visual features (which are both perceptually salient and visually distinctive to be detected robustly, see Shi and Tomasi 1994; Knapek et al. 2000) to estimate the changes in self-position and, from this

information, to build a local map of the environment. In robotics, this technique, termed SLAM (Simultaneous localization and mapping Thrun 2002), usually requires an independent set of sensors to compute the robot's motion over time. While inertial odometry can be quite accurate for wheeled robots and vehicles where the angles of steering and wheel rotation encoders provide reliable estimates of the motion from a given starting location, it is much more complex in the case of pedestrian displacement. The motion of a walking person exhibits high variation in velocity and trajectory, and the estimate of the number of steps, step length, and heading from pedometers and accelerometers is rarely of sufficient accuracy. In addition, visual odometry as well as inertial odometry inevitably accumulate errors over time, which results in an increasing drift in position with time, if these errors are not regularly corrected with an absolute reference position. For these reasons, the proposed system employs visual landmarks with known geographic positions (annotated in the GIS database, see Sect. 3.4) to refine the pedestrian position estimated by a GPS. Few other systems have proposed such a solution (see Park et al. 2009).

The vision unit of the NAVIG system is designed to extract relevant visual information from the environment in order to compensate for the visual deficit of the user. To do so, the user is equipped with head-mounted stereoscopic cameras providing images of the surroundings, which are processed in real-time by computer vision algorithms localizing objects of interest. Different categories of targets can be detected depending on the circumstances. These can be either objects requested by the user, or geolocated landmarks requested by the system and used to compute the user's position. In both cases, the core function of performing pattern-matching remains the same, relying on a biologically inspired image processing library called SpikeNet.[2]

The SpikeNet recognition algorithms are based on biological vision research, specifically on the mechanisms involved in ultra-rapid recognition within the human visual system (Thorpe et al. 1996). Computational models of these mechanisms led to the development of a recognition engine providing very fast processing, high robustness to noise and light conditions, and good invariance properties to scale or rotation. SpikeNet uses a model for each visual pattern that needs to be recognized, which encodes the visual saliencies of the shape in a small structure (30×30 px (pixel) patch), thus requiring very little memory. Several models are usually needed to detect a specific object from different points of view, but even in an embedded system millions of models can easily be stored and loaded when needed given their small size. In terms of speed, the

---

[2] SpikeNet Technology, http://www.spikenet-technology.com.

Fig. 3 Examples of geolocated visual landmarks used for user-positioning (shop sign, facade, road sign, mailbox)

processing time strongly depends on the size of the input and target images. As an example, for the current NAVIG prototype based on a notebook equipped with an Intel i7 820QM processor (1.73 GHz) and 4 GB memory, the recognition engine achieved a stable analysis speed of 15 fps (frames per second) with a 320×240 px image stream while concurrently searching for 750 visual shapes of size 120×120 px. Maintaining a rate of 15 fps, the number of different models that could be tested was reduced to 250 for 60×60 px models, and to 65 with 30×30 px models.

In the "object-identification" mode, the user expressly requests an object of current interest. Then the system automatically loads the corresponding models. When the target is detected, the relative location is computed by stereovision methods. The object is then rendered via virtual 3D sounds (see Sect. 3.5), with its position updated every 60 ms and interpolated between updates using head rotation data provided by an inertial measurement unit attached to the helmet. This function plays a very important role in the NAVIG device as it restores a functional visuomotor loop (see Sect. 4) allowing a visually impaired user to move his/her body or hand to targets of interest (Dramas et al. 2010).

The second function of the vision unit, detailed in (Brilhault et al. 2011), provides user-positioning features. As the user is guided to a chosen destination, the system attempts to detect geolocated visual targets along the itinerary. These detected objects are not displayed to the user but are used as anchors to refine the current GPS position. These visual landmarks (termed visual reference points, *VP*) can be objects such as signs, statues, mailboxes as well as facades, particular layouts of buildings, etc. (see Fig. 3), which are stored in the Geographic Information System with their geographic coordinates. When a visual landmark is detected, the user's position can be estimated using the distance and angle of the target relative to the cameras (provided by the stereoscopic depth map) and data from other sensors (e.g., magnetometers, accelerometers). This positioning method can provide an estimate relying exclusively on vision, in the event of GPS signal loss, or can be integrated into a larger fusion scheme as described in Sect. 3.3. It should be emphasized that it is not the user who determines the particular set of visual targets to be loaded. Instead, the system automatically loads the models corresponding to the rough location of the user given by the GPS.

The creation of models remains a key issue, which in the current version of the system has been performed manually

from recorded videos of the evaluation test site. For generic objects that might be requested, the device could be preconfigured with an initial set of models covering a large number of common objects, built automatically from segmented image databases. A tool providing this automatic generation of models is currently under development. Each individual user should also be able to add new models. Model creation could entail the user rotating the new unknown object in front of the cameras, allowing segmentation of the image of interest within the optical flow, and then dynamic creation of an ensemble of models to ensure adequate coverage of the object.

For visual landmarks, the problem is more complex as the correct GPS coordinates of the target are required. Exploitation of services similar to Google Street View[3] could allow for automatic construction of model sets in a new area. To examine this approach, the Topographic Department of the City of Toulouse has contributed 3D visual recordings of all streets of the city to the project, combining eight different views taken from car-mounted cameras at one meter intervals, combined with laser data allowing retrieval of GPS coordinates of any point in these images. With this database, it could be possible to randomly search for patterns throughout the city streets which are distinctive enough so as not to trigger false detections in neighboring streets, and to automatically store them as visual reference points with their associated coordinates.

Some studies have presented interesting and complementary approaches based on social cooperation (Völkel et al. 2008). It is proposed that the annotation of SIG databases may rely on data collected "on the move" by users themselves, combined with information gathered by the internal sensors of the device (e.g., compass, GPS). To increase data sources and facilitate sharing between users, a client-server architecture was proposed with the database stored on a remote server. The database is constantly updated and anonymously shared among users. An alternate method to add visual reference points could be based on the cooperate effort of "web workers." One such approach, VizWiz,[4] combines automatic and human-powered services to answer visual questions for visually impaired users.

3.3 Data fusion for pedestrian navigation

The measurement of physical quantities such as position, orientation, and acceleration relies on sensors that inherently report approximated values. This fact, in addition to occasional sensor dropout or failure, results in a given system receiving somewhat inaccurate or incomplete

information. As such, the NAVIG system employs a collection of different sensors to obtain the same information, for example position. These different estimates must then be combined to provide the best estimate through a data fusion model. There are three main issues identified in sensor data fusion:

- Interpretation and representation: Typically handled with probabilistic descriptions (Durrant-Whyte 1988).
- Fusion and estimation: Methods such as Bayesian estimation (Berger 1985) and Kalman filtering (Bar-Shalom 1987) are widely used.
- Sensor management: Solutions are based either on a centralized or a decentralized sensor architecture (Mitchell 2007).

Centralizing the fusion process combines all of the raw data from all the sensors in one main processing module. In principle, this is the best way to realize the data fusion, as all the information is still present. In practice, centralized fusion frequently overloads the processing unit with large amounts of data. Preprocessing the data at each sensor drastically reduces the required data flow, and in practice, the optimal setup is usually a hybrid of these two types. Care must be taken in fusing different types of data, ensuring that transformations are performed to provide a unified coordinate system before the data fusion process.

It is important to note that different sensor systems operate with different and sometimes variable refresh rates. The sensor fusion strategy takes into account the amount of time from the last received data to automatically adjust the weights (i.e. estimated accuracy) attributed to each sensor. For example, a sudden drop-out in GPS signal for more than a few seconds (in an urban canyon) would gradually reduce the weight attributed to the GPS data. Similar corrections would be applied to the other sensors, depending on the specific time characteristics of each of these sensors.

Several GPS systems equipped with different sensors have been developed to increase navigation accuracy in vehicles (Cappelle et al. 2010). In these systems, the inertia of the vehicle is important and dead reckoning strategies are appropriate for predicting the position at the next time step (p_{t+1}). Furthermore, velocity and trajectory of a vehicle exhibit smooth and relatively slow variations. Finally, there is a high probability that the vehicle follows the sense of traffic known for the given side of the road. All these elements make accurate position estimation possible for vehicles, but they are not applicable in the case of pedestrian navigation.

The aim of the current fusion algorithm is to manage these two issues: first, by taking into account pedestrian ways of moving; secondly, by employing user mounted cameras that will recognize natural objects in urban scenes and allow for a precise estimate of the position of the user.

Fig. 4 Process of computing the user location when a *VP* (here the bench with known GPS coordinates) is detected by the embedded vision module. Using the 3D position of the object in the camera reference frame (obtained through stereovision), and the orientation of the head (acquired by an Inertial Measurement Unit), it is possible to estimate the geolocation of the user
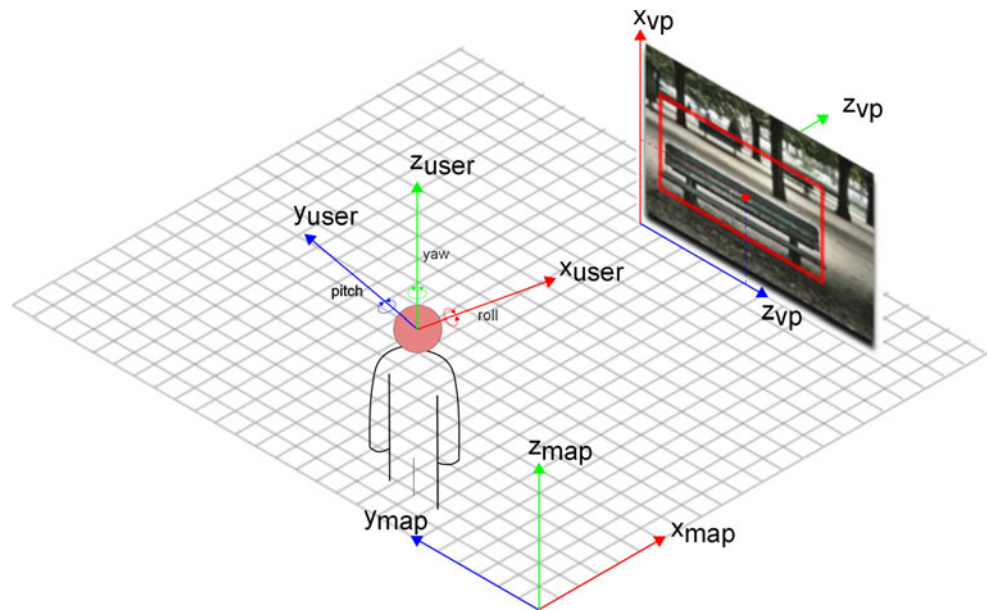
This solution avoids the time and expense of equipping the environment with specific instrumentation as mentioned in Park et al. (2009), Bentzen and Mitchell (1995).

The fusion of positional information from the image recognition and geolocalization systems in real-time is a novel approach that results in an improvement in precision for the estimation of the user's position. The approach is to combine satellite data from the Global Navigation Satellite System (GNSS) element and position estimations based on the visual reference points with known geographic coordinates (see Fig. 4). Using a detailed database that contains embedded coordinates of these landmarks, the position of the user can be geometrically estimated to a high degree of precision. The integration of accelerometers provides added stability in separating tracking jitter from actual user motion.

The fusion algorithm uses three different inputs. First, a commercial GPS sensor, assisted by an inertial system, provides accurate coordinates. Second, a GIS is used to verify that positions are coherent with map constraints. Finally, the vision system provides information on any geographically located objects detected. Figure 5 presents preliminary results for the estimation of user location by the fusion of information provided by the GPS receiver and the location estimate relying on embedded vision (Brilhault et al. 2011).

3.4 Geographical information system

A Geographic Information System (GIS) has been defined by Burrough (1986) as a tool for capturing, manipulating, displaying, querying, and analyzing geographic data. The GIS is an important component in the design of an *electronic orientation aid* for VI persons (Golledge et al. 1998). Relying on a digitized spatial database and analytical tools, the NAVIG GIS module provides the user with accurate environmental information to ensure the success of the navigation task.

3.4.1 Digitized spatial data base

Many studies (see e.g. Fletcher 1980) have shown that building a cognitive map is useful to resolving spatial tasks. To use GIS databases for VI pedestrian navigation aids, it is necessary to augment them with additional classes of important objects, in order to provide the user with important specific information concerning the itinerary and surroundings (see Jacobson and Kitchin 1997). This information must then be rendered during preparatory planning or actual navigation and may serve to build sparse but useful representations of the environment.

In the context of wayfinding aids for blind pedestrians, (Gaunet and Briffault 2005) showed that adapted GIS databases for pedestrian navigation should include streets, sidewalks, crosswalks, and intersections. In addition, they specified that guidance functions consist of a combination of orientation and localization, goal location, intersection, crosswalks, and warning information as well as of progressions, crossings, orientations, and route-ending instructions. Therefore, all of those stated features concerning the path, the surroundings, and adapted guidance should be collected and stored with a high degree of spatial precision in the GIS. They should be incorporated into route selection procedures and displayed to the user during on-site guidance. Their utility during preliminary preparation of a journey should also be examined.

Fig. 5 A test path at the Toulouse University campus indicating buildings (*gray polygons*) and archways (*pink polygons*). *VP* with known GPS coordinates detected during the journey are shown with *green dot*. The different paths shown include the expected itinerary (*purple filled square*), commercial GPS sensor positioning (*yellow diamond*), user locations estimated by the vision module (*red star*), and the more accurate position computed using the data fusion module (*blue dot*). See the realization of a variety Sect. 3.4.1 for label descriptions
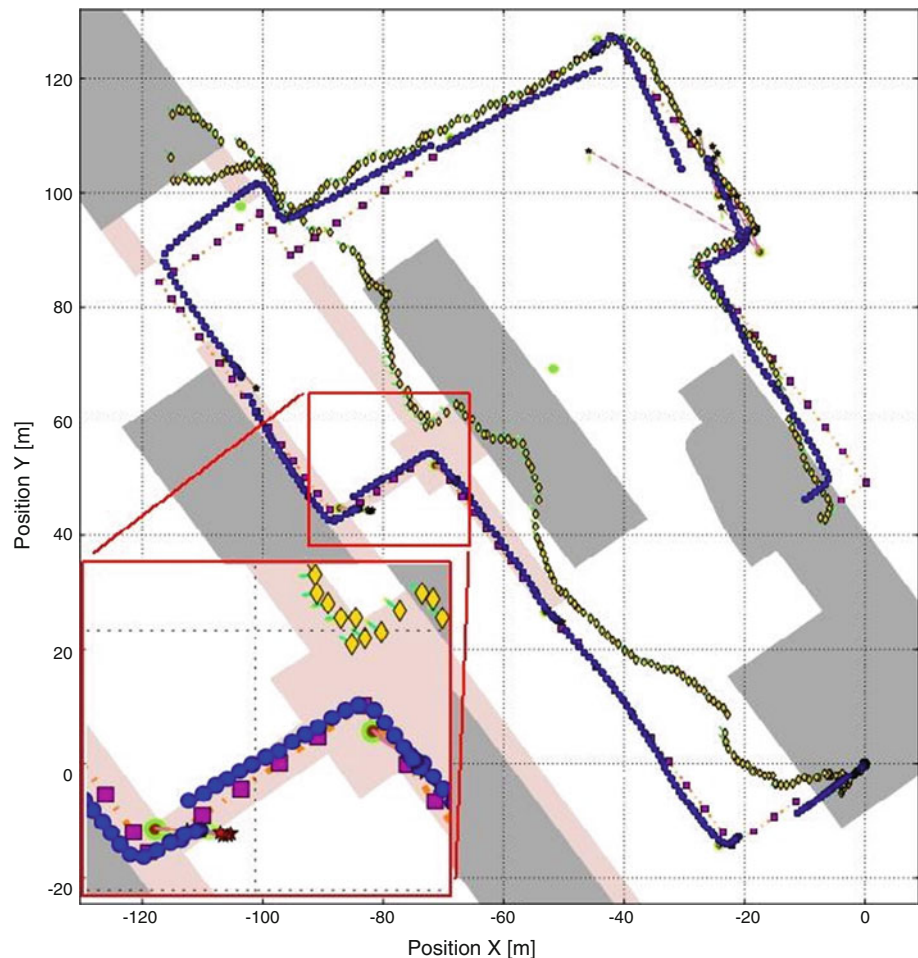
Currently, commercial GIS systems have been almost exclusively developed for car travel. A series of brainstorming sessions and interviews were conducted with potential VI users and orientation and mobility (O&M) instructors (see Sect. 2). The results led to five classes of objects that should be included in a GIS adapted to VI pedestrian navigation:

1. Walking Areas (*WA*): All possible pedestrian paths as defined in Zheng et al. (2009) (e.g., sidewalks, and pedestrian crossings).
2. Landmarks (*LM*): Places or objects that can be detected by the user in order to make a decision or confirm his own position along the itinerary (e.g. changes in texture of the ground, telephone poles, or traffic lights).
3. Difficult Points (*DP*): Places that represent potential mobility difficulties for VI pedestrians (see Sect. 2.2).
4. Points of Interest (*POI*): Places that are potential destinations or that contain interesting features. When they are not used as a destination, they are useful or interesting places offering the user a better understanding of the environment while traveling (e.g. public buildings, shops, etc.).

5. Visual Reference Points (*VP*): Geolocalized objects used by the vision module.

For each object in the database, multiple tags were possible. For instance, a bus stop was tagged as a *LM* because it can be detected by the user. It was also tagged as *POI* as it is a potential destination and as a *VP* if it could be detected by the artificial vision module. In addition, the user has the possibility to add specific locations (such as home, work, or sidewalks which could be slippery when wet) that will be integrated in a specific user layer of the GIS. The class of these objects is called Favorite Point (*FP*). Each point will be associated with a specific tag defined by the user.

3.4.2 GIS software: route selection

Once the user position and destination have been determined, route selection is necessary and is usually included in the GIS component. It is defined as the procedure of choosing an optimal pathway between origin and destination. When considering pedestrian navigation, the shortest path might be appropriate but should rely on a GIS database including essential information for pedestrian

mobility (e.g., sidewalks and pedestrian crossings). Traditionally, route or path selection is assumed to be the result of minimization procedures such as selecting the shortest or the quickest path. For visually impaired users, a longer route may be more convenient than a shorter route, in order to avoid various obstacles or other difficulties. These route optimization rules can vary between individual users, due to mobility training or experience and other personal factors (see Sect. 2.2). An adapted routing algorithm for visually impaired pedestrians has been proposed to improve path choice (Kammoun et al. 2010). The aim is to find the preferred route that connects the origin and destination points. The selected path is represented as a road map containing a succession of Itinerary Points (*IP*) and possibly Difficult Points (*DP*), such as pedestrian crossing and intersections, linked by *WA* as well as a collection of nearby *POI*, *FP*, *LM*, and *VP* as defined in Sect. 3.4.1.

3.5 Spatial audio

While the locations of the user and obstacles, and the determination of the itinerary to follow to attain the intended goal, are fundamental properties of the system, this information is not useful if it cannot be exploited by the user. The NAVIG system proposes to make use of the human capacity for hearing, and specifically spatial audition, by presenting guidance and navigational information via binaural 3D audio scenes (Begault 1994; Loomis et al. 1998). The 3D sound module provides binaural rendering over headphones using a high performance spatialization engine (Katz et al. 2010) developed under the Max/MSP programming environment.[5]

In contrast to traditional devices that rely on turn by turn instructions, the NAVIG consortium is working toward providing spatial information to the user concerning the trajectory, their position in it, and important landmarks. This additional information will help users become more confident when navigating in unknown environments.

Many visually impaired persons are already exploiting their sense of hearing beyond the capacities of most sighted people. Using the auditory modality channel to provide additional, and potentially important, information requires careful design in order to minimize cognitive load and to maximize understanding. Although the use of stereo headphones is required to produce binaural 3D sound, wearing traditional headphones results in a certain degree of masking of real sounds, which is problematic for VI individuals. Instead, a solution employing bonephones—headphones that work via the transmission of vibrations against the side of head, transmitting the sound via bone

conduction, was adopted. Previous studies have demonstrated the efficient use of bonephones within a virtual 3D audio orientation context (Walker and Lindsay 2005). These particular headphones, situated just in front of the ears without any obstruction of the ear canal or pinna, permit the use of 3D audio without any masking of the real acoustic environment. Because of the bonephone's complex frequency response, tailored equalization is necessary in order to properly render all the spectral cues of the Head Related Transfer Function.

There are many instances where textural verbal communication is optimal, such as indicating street names or landmarks. At the same time, a path or a spatial display are not verbal objects, but spatial objects. As such, the exploitation of auditory trajectory rendering can be more informative and more intuitive than a list of verbal instructions. The ability to have a global representation of the surroundings (survey representation) and a sense of the trajectory (route representation) is also highly desirable.

In contrast to previous works on sensory substitution, where images captured by a camera are directly transformed into sound, the aim of the 3D sound module is to generate informational auditory content at the spatial position, which directly coincides with that of a specific target. Various methods for semantic or informative spatial sonification have been shown to be effective in spatial manipulations of virtual objects (Férey et al. 2009) and scientific exploration and navigation tasks within large abstract datasets (Vézien et al. 2009) in multimodal virtual environments. Spatial sonification for spatial data exploration and guidance (Katz et al. 2008) and target acquisition (Ménélas et al. 2010) in audio and audio-haptic virtual environments without visual renderings have also been shown to be effective in previous studies.

A previous study has examined the precision of hand reaching movement toward nearby real acoustic sources through a localization accuracy task (Dramas et al. 2008). Results showed that the accuracy of localization varies relative to source stimuli, azimuth, and distance. Taking into account these results, preparations are underway for a grasping task experiment with virtual sounds and different stimuli to optimize the accuracy of localization.

3.6 User guidance

Once the user position has been determined, and the location or object identified, the primary task for the assistive system is to guide the user in a safe and reliable manner. Depending on their preferences and knowledge of the system, users have the possibility to choose different levels of detail of information and the way that this information will be presented to them. A series of brainstorming sessions and interviews with VI panel members

---

[5] Max visual programming language for music and multimedia (2012) http://cycling74.com/products/maxmspjitter/.

identified at least two types of navigation to be considered (Brunet 2010).

First, the *normal* mode is used for point-to-point guidance along a calculated itinerary from point A to a point B. In this mode, the user needs only a minimum amount of information to understand and perform the navigation task, with only the *IP*, *DP*, and *LM* elements being necessary. In contrast, in *exploration* mode, the user is interested in exploring a neighborhood or a specific itinerary. As such, there is a need to provide additional information, such as the presence and location of bakeries, municipal buildings, or bus stops. This mode requires the presentation of *IP*, *DP*, *LM*, and *POI*. At each use, the user can personalize the presented information by selecting certain types of *POI* or *LM* that are of personal interest. To facilitate this categorical presentation, each object class is divided into several categories that can be used to filter the information (7 categories of *POI*, 4 *LM*, and 3 *FP*).

Different levels of verbalization are provided in the NAVIG system depending on user needs (see Sect. 2). If the *IP* are always rendered by placing a virtual 3D sound object at the next waypoint along the trajectory, the *POI*, *PF*, and *LM* can be rendered using text-to-speech (TTS) or semantic sounds. All presented information is spatialized so that the user hears the description of each object coming from its corresponding position. Users can choose to use only TTS, a mix of spatialized TTS and semantic sounds, or only semantic sound. To accommodate spatialized TTS, a version of Acapela[6] was incorporated into the 3D real-time rendering system (Katz et al. 2010).

For the semantic sound mode, it is important to study how all the information will be displayed in an ergonomic and intuitive sound display. For a navigation task using virtual audio display with several types of information (*IP*, *DP*, *POI*, . . .), two types of auditory cues can be used to create beacon sounds:

- Auditory Icons are described in Gaver (1986) as "everyday sounds mapped to computer events by analogy with everyday sound producing events". They are brief sounds that can be seen as the auditory equivalent of visual icons used in personal computer.
- Earcons are abstract, synthetic, and mostly musical tones or sound patterns that can be combined in a structured way to produce a sound grammar. They are defined in Blattner et al. (1989) as "non-verbal audio messages used in the computer interface to provide information to the user about some computer objects, operation or interaction". Earcons allow for the construction of a syntactic hierarchical system in order to represent data trees with several levels of information.

- Spearcons, introduced in Walker et al. (2006), use spoken phrases sped up until they may no longer be recognized as speech. Built on the basis of the text describing the information they represent, spearcons can easily be created using TTS software and an algorithm to speed up the phrase. Since the mapping between a spearcon and the object it represents is non-arbitrary, only a short training is required.

The advantages and disadvantages of these various sonification methods are relatively well known in auditory displays. Several studies have explored the learnability of such displays; for example, (Dingler et al. 2008) presents the superiority of spearcons compared to auditory icons and earcons in term of learnability. Other studies explored navigation performance, comparing different types of beacon sounds and the effects of the display rate (see Walker and Lindsay 2006; Tran et al. 2000 for an ergonomic evaluation of acoustic beacon characteristics and differences between speech and sound beacon). While the effectiveness and the efficiency of acoustic beacons have been well investigated (see Loomis et al. 1994, 1998, 2005, 2006 for systematic studies of the value of virtual sound for guidance), studies concerning user satisfaction with auditory navigation systems are still severely lacking. For the NAVIG project, the concept of morphological earcons (morphocons) has been introduced in order to improve user satisfaction through the development of a customizable user audio-interface.

Morphocons (morphological earcons) allow the construction of a hierarchical sound grammar based on temporal variation of several acoustical parameters. With this method, it is then possible to apply these morphological variations to all types of sounds (natural or artificial) and therefore to construct an infinite number of sound palettes while maintaining a certain level of coherence among objects or messages to be displayed. For the NAVIG project, a semantic sound grammar has been developed to allow the user to rapidly identify and differentiate between each class of objects (*IP*, *DP*, *POI*, *PF*, and *LM*) and to be informed about the subcategories within each class. This grammar has been established so that each sound can be easily localized (i.e. large spectrum, sharp attack), the possibility of confusion between classes is minimized, and considerations are made concerning the superposition of the virtual soundscape and the real acoustic world. The semantic sound grammar is illustrated in Fig. 6 and is described as follows:

- *IP* : a brief sound
- *DP* : a sequence of two brief sounds
- *LM* : a rhythmic pattern of three brief sounds. Rhythmic variations of this pattern allow for the differentiation of *LM* type.

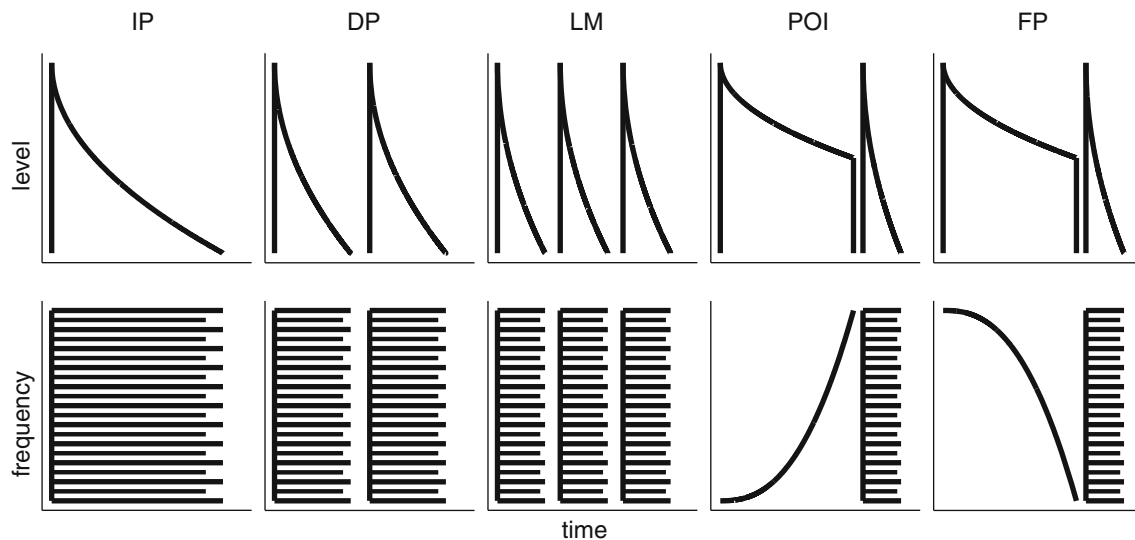[6] Acapela group, http://www.acapela-group.fr/.

Fig. 6 Illustration of semantic sound grammar indicating (*upper*) intensity and (*lower*) frequency profiles of each element of the palette

- *POI* : a sound whose frequency increases steadily, followed by a brief sound. The first sound is common to all categories of *POI*, while the brief sound differentiates between them.
- *FP* : a sound whose frequency decreases steadily, followed by a brief sound. The first sound is common to all categories of *FP*, while the brief sound differentiates between them.

Sound durations are between 0.2 and 1.5 s. This common grammar allows for the realization of a variety of sound palettes (e.g., birds, water, musical, videogame) satisfying individual user preferences in terms of sound esthetic while maintaining a common semantic language. As such, switching between palettes should not imply any significant change in cognitive load or learning period. Three different sound palettes (*natural*, *instrumental*, and *electronic*)[7] were constructed and perceptually evaluated by 60 subjects (31 sighted and 29 blind subjects) with an online classification test.[8] Results showed a good recognition rate for discrimination between the categories (78 ± 22 %) with no difference between sighted and blind subjects. Concerning the discrimination between subcategories, the recognition rate was 63 ± 23 % for the *POI*, 58 ± 29 % for the *LM* and 87 ± 19 % for the *FP*. These results showed that the rhythm variations used to differentiate the *LM* subcategories were too similar and should be improved. They also pointed to specific problems for some sounds within each palette which were problematic. On the basis of these results, three new sound palettes are being created for the next phase of navigation testing. Additional details concerning the developed morphocons can be found in Parseihian and Katz (2012).

4 Near-field assistive mode

In its simplest form, direct point-to-point guidance can be used to attain any requested object. The task of object localization or grasping then consists of a direct loop between the recognition algorithm detecting the target and the sound spatialization engine attributing and rendering a sound object at the physical location of the actual object. As such, the architecture for near-field guidance is dynamically simplified in an attempt to optimize performance and minimize system latencies.

When the object of interest is detected, the position of the target is directly sent to the sonification module. Rapid image recognition of objects in the camera's field of view provides head-centered coordinates for detected objects, offering built-in head tracking (see Fig. 7). For robustness in the case of lost identification or objects drifting out of the field of view, a 3D head orientation tracking device is also included to interpolate object positions, insuring fluidity and maintaining a refresh latency of no more than 10 ms with respect to head movements.

To improve the micro-navigation task, route selection should also be addressed. Unlike in the macro-scale pedestrian navigation, trajectory determination in the near-field, or in indoor situations, is more difficult, as the only data source is that from the image recognition platform. Nevertheless, intelligent navigation paths could be developed even in these situations. A contextual example of a typical situation would be to find a knife on a cluttered kitchen counter top. The user requests the knife. While the

---

Fig. 7 The user interacts with the system via voice recognition; in this case, he would like to grab his phone, the system then activates the research of the corresponding models in the artificial vision module. If the object is detected, the position of the target is directly sent to the sonification module to display the position of the phone

object can be easily and quickly identified, and its position determined, in this context there can be a number of obstacles which could be in the direct path to the knife, such as seasoning bottles. In addition, a knife has a preferred orientation for grasping, and it would be preferable if the assistive device was aware of the orientation of the object and would direct the user accordingly to the handle, and not the blade.

Outside the context of micro-navigation, this assistive device may also serve for general object recognition. Indeed, during the user-centered design sessions, participants mentioned the recurrent problem of distinguishing

among similar objects (e.g., canned foods, bank notes). The NAVIG prototype has been tested in a study where participants had to classify different euro (€) currency notes (Parlouar et al. 2009). As there are few mobile systems that are able to satisfactorily recognize different bank notes (see Liu 2008), the aim was to evaluate this sub-function of the device's vision module. Due to the high-speed and robustness of the recognition algorithm, users were able to identify 100 % of the bills that were presented and performed the sorting task flawlessly. Average measured response times (including bill manipulation, recognition, and classification tasks) were slightly above 10 s per bill. Users were in agreement that the usability of the system was good.

5 NAVIG guidance prototype

The first functional prototype (shown in Fig. 8) operates on a laptop. The artificial vision module currently uses video streams from two head-mounted cameras (320 × 240 px at 48 Hz). The prototype employs a stereo camera pair with an approximately 100° viewing angle, allowing for the computation of distance to the objects based on stereoscopic disparity and the calibration matrix of the lenses. The prototype hardware is based on an ANGEO GPS (NAVOCAP, Inc), a BumbleBee stereoscopic camera system (Point Grey Research, Inc), an XSens orientation tracker, headphones, microphone, and a notebook computer. The NAVIG prototype has been successfully tested on a simple scenario in the Toulouse University campus.

Fig. 8 NAVIG Prototype V1

Preliminary experiments with this prototype have shown that it is possible to design a wearable device that can provide fully analyzed information to the user.

The design of an assistive device for visually impaired users must take into account users' needs as well as their behavioral and cognitive abilities in spatial and navigational tasks. This first prototype, pretested by blindfolded participants, will be evaluated in the fall of 2012 by a panel of 20 visually impaired participants involved in the project, with the headphones replaced by bonephones.

6 Conclusion

This paper has introduced the NAVIG augmented reality assistance system for the visually impaired whose aim is to increase individual autonomy and mobility in the context of both sensing the immediate environment and pedestrian navigation. Combining satellite, image, and other sensor information, high precision geolocalization is achieved. Exploiting a rapid image recognition platform and spatial audio rendering, detailed trajectories can be determined and presented to the user for attaining macro- or micro-navigational destinations. An advanced dialog controller is being developed to facilitate usage and optimize performance for visually impaired users.

This kind of assistive device, or *electronic orientation aid*, does not replace traditional mobility aids such as the cane or the guide dog, but is considered as an additional device providing the VI user with important information for spatial cognition including landmarks (e.g., important points on the itinerary related to decision or confirmation), routes to follows (guidance), and spatial description. Specifically, it restores fundamental visuomotor processes such as grasping, heading, and piloting. In addition, it allows selection of adapted routes for VI pedestrians. Finally, we suggest that a spatial environment description mode, based on 3D synthesis of the relative location of important points in the surrounding, may help visually impaired users generate a sparse but functional mental map of the environment. This function will be evaluated as part of the ongoing ergonomic evaluations of the NAVIG system.

References

Afonso A, Blum A, Katz BFG, Tarroux P, Borst G, Denis M (2010) Structural properties of spatial representations in blind people: scanning images constructed from haptic exploration or from locomotion in a 3-D audio virtual environment. Memory Cogn 38:591–604

Allen GL (2000) Principles and practices for communicating route knowledge. Appl Cogn Psychol 14:333–359

Auvray M, Myin E (2009) Perception with compensatory devices. From sensory substitution to sensorimotor extension. Cogn Sci 33:1036–1058

Auvray M, Hanneton S, O'Regan JK (2007) Learning to perceive with a visuo-auditory substitution system: localization and object recognition with the voice. Perception 36:416–430

Bar-Shalom Y (1987) Tracking and data association. Academic Press Professional, ISBN: 0-120-79760-7

Begault DR (1994) 3-D sound for virtual reality and multimedia. Academic Press, Cambridge

Bentzen B, Mitchell P (1995) Audible signage as a wayfinding aid: Comparison of Verbal Landmarks® and Talking Signs®. J Vis Impair Blind 89:494–505

Berger JO (1985) Statistical decision theory and bayesian analysis (2nd edn). Springer Series, ISBN: 978-0387960982

Bisseret A, Sebillote S, Falzon P (1999) Techniques pratiques pour l'étude des activités expertes. Octarès Editions, Toulouse

Blattner MM, Sumikawa DA, Greenberg RM (1989) Earcons and icons: their structure and common design principles. SIGCHI Bull 21:123–124

Brilhault A, Kammoun S, Gutierrez O, Truillet P, Jouffrais C (2011) Fusion of artificial vision and GPS to improve blind pedestrian positioning. International conference on new technologies, mobility and security, IEEE, France

Brunet L (2010) Étude des besoins et des stratégies des personnes non-voyantes lors de la navigation pour la conception d'un dispositif d'aide performant et accepté (Needs and strategy study of blind people during navigation for the design of a functional and accepted aid device). Master's thesis, Department of Ergonomics, Université Paris-Sud, Orsay, France

Buisson M, Bustico A, Chatty S, Colin F-R, Jestin Y, Maury S, Mertz C, Truillet P (2002) Ivy: un bus logiciel au service du développement de prototypes de systèmes interactifs. 14th French-speaking conference on human–computer interaction (IHM '02) pp 223-2 26

Burrough PA (1986) Principles of geographical information systems for land resources, assessment. Oxford, Clarendon Press, Monographs on soil and resource surveys, No. 12

Canadian Institute for the Blind (2005) Inégalité des chances : Rapport sur les besoins des personnes aveugles ou handicapées visuelles vivant au Canada. Technical Report. http://www.cnib.ca/fr/apropos/publications/recherche

Cappelle C, El Najjar ME, Pomorski D, Charpillet F (2010) Multi-sensors data fusion using dynamic bayesian network for robotised vehicle geo-localisation. International conference on information fusion, June 30–July 3 2008, Cologne, pp 1–8

Denis M (1997) The description of routes: acognitive approach to the production of spatial discourse. Curr Pychol Cogn 16:409–458

Dingler T, Lindsay J, Walker BN (2008) Learnability of sound cues for environmental features: Auditory icons, earcons, spearcons, and speech, methods, pp 1–6

Dramas F, Oriola B, Katz BFG, Thorpe S, Jouffrais C (2008) Designing an assistive device for the blind based on object localization and augmented auditory reality. ACM conference on computers and accessibility, ASSETS, Halifax, Canada, pp 263–264

Dramas F, Thorpe SJ, Jouffrais C (2010) Artificial vision for the blind: a bio-inspired algorithm for objects and obstacles detection. J Image Graph 10(4):531–544

Durrant-Whyte HF (1988) Sensor models and multisensory integration. Special issue on Sensor Data Fusion, ISSN: 0278-3649 7(6):97–113

Férey N, Nelson J, Martin C, Picinali L, Bouyer G, Tek A, Bourdot P, Burkhardt JM, Katz BFG, Ammi M, Etchebest C, Autin L (2009) Multisensory VR interaction for protein-docking in the CoRS-AIRe project. Virtual Reality 13:273–293

Fletcher JF (1980) Spatial representation in blind children. 1: development compared to sighted children. J Vis Impair Blind 381–385

Gallay M, Denis M, Auvray M (2012) Navigation assistance for blind pedestrians: guidelines for the design of devices and implications for spatial cognition. In: Thora Tenbrink, Jan Wiener, Christophe Claramunt (eds) Representing space in cognition: Interrelations of behaviour, language, and formal models. Oxford University Press, UK (in press)

Gaunet F, Briffault X (2005) Exploring the functional specifications of a localized wayfinding verbal aid for blind pedestrians: simple and structured urban areas. Hum Comp Interact 20:267–314

Gaver W (1986) Auditory icons: using sound in computer interfaces. Hum-Comput Interact 2:167–177

Golledge RG, Klatzky RL, Loomis JM, Speigle J, Tietz J (1998) A geographical information system for a GPS-based personal guidance system. Int J Geograph Inform Sci 727–749

Golledge RG, Marston JR, Loomis JM, Klatzky RL (2004) Stated preferences for components of a Personal Guidance System for nonvisual navigation. J Vis Impair Blind 98:135–147

Helal A, Moore SE, Ramachandran B (2001) Drishti: an integrated navigation system for visually impaired and disabled. International symposium on wearable computers ISWC, IEEE Computer Society, Washington DC, pp 149–156

Hub A, Diepstraten J, Ertl T (2004) Design and development of an indoor navigation and object identification system for the blind. ACM SIGACCESS accessibility and computing (77–78):147–152

Jacobson RD, Kitchin RM (1997) GIS and people with visual impairments or blindness: Exploring the potential for education, orientation, and navigation. Trans Geograph Inform Syst 2(4):315–332

Kammoun S, Dramas F, Oriola B, Jouffrais C (2010) Route Selection Algorithm for Blind Pedestrian. International conference on control, automation and systems, IEEE, KINTEX, Gyeonggi-do, Korea, pp 2223–2228

Kammoun S, Parseihian G, Gutierrez O, Brilhault A, Serpa A, Raynal M, Oriola B, Macé M, Auvray M, Denis M, Thorpe S, Truillet P, Katz BFG, Jouffrais C (2012) Navigation and space perception assistance for the visually impaired: the NAVIG project. Ingénierie et Recherche Biomédicale 33:182–189

Katz BFG, Rio E, Picinali L, Warusfel O (2008) The effect of spatialization in a data sonification exploration task. In: Proceedings of 14th meeting of the international conference on auditory display (ICAD), Paris 24–27 June, pp 1-7

Katz BFG, Truillet P, Thorpe S, Jouffrais C (2010) NAVIG: Navigation assisted by artificial vision and GNSS. Workshop on multimodal location based techniques for extreme navigation (Pervasive 2010), Helsinki

Katz BFG, Rio E, Picinali L (2010) LIMSI spatialisation engine, InterDeposit Digital Number IDDN.FR. 001.340014.000.S.P. 2010.000.31235

Katz BFG, Dramas F, Parseihian G, Gutierrez O, Kammoun S, Brilhault A, Brunet L, Gallay M, Oriola B, Auvray M, Truillet P, Denis M, Thorpe S, Jouffrais C (2012) NAVIG: guidance system for the visually impaired using virtual augmented reality. J Technol Disability 24(2) (in press)

Klatzky RL, Marston JR, Giudice NA, Golledge RG, Loomis JM (2006) Cognitive load of navigating without vision when guided by virtual sound versus spatial language. J Exp Psychol Appl 12:223–232

Knapek M, Oropeza RS, Kriegman DJ (2000) Selecting promising landmarks, Proc. ICRA'00. IEEE international conference on robotics and automation vol 4, pp 3771–3777

Liu X (2008) A camera phone based currency reader for the visually impaired. In: Proceedings of the 10th international ACM SIGACCESS conference on computers and accessibility, Halifax, Nova Scotia, Canada, pp 305–306

Loomis JM, Golledge RG, Klatzlty RL, Speigle J, Tietz J (1994) Personal guidance system for visually impaired. In: Proceedings of first annual ACM conference on assistive technologies (Assets '94), pp 85–90

Loomis JM, Golledge RG, Klatzky RL (1998) Navigation system for the blind: auditory display modes and guidance. Presence: Teleoperators Virtual Environ 7:193–203

Loomis JM, Marston JR, Golledge RG, Klatzky RL (2005) Personal guidance system for people with visual impairment: a comparison of spatial displays for route guidance. J Vis Impair Blind 99:219–232

Loomis JM, Golledge RG, Klatzky RL, Marston JR (2006) Assisting wayfinding in visually impaired travelers. In: Allen G (eds) Applied spatial cognition: from research to cognitive technology, Lawrence Erlbaum Associates, Mahwah

Marston JR, Loomis JM, Klatzky RL, Golledge RG (2006) Smith EL Evaluation of spatial displays for navigation without sight. ACM Trans Appl Percept 3:110–124

Ménélas B, Picinali L, Katz BFG, Bourdot P (2010) Audio haptic feedbacks in a task of targets acquisition. IEEE symposium on 3d user interfaces (3DUI 2010), Waltham, USA, pp 51–54

Mitchell HB (2007) Multi-sensor data fusion: an introduction. Springer, ISBN: 978-3540714637

Noordzij ML, Zuidhoek S, Postma A (2006) The influence of visual experience on the ability to form spatial mental models based on route and survey descriptions. Cognition 100:321–342

Park S-K, Suh YS, Do TN (2009) The pedestrian navigation system using vision and inertial sensors. ICROS-SICE international joint conference on 2009, Fukuoka, Japan 18–21 Aug., pp 3970–3974

Parlouar RD, Macé FM, Jouffrais C (2009) Assistive device for the blind based on object recognition: an application to identify currency bills. ACM conference on computers and accessibility (ASSETS 2009), Pittsburgh, PA, USA, pp 227–228

Parseihian G, Katz BFG (2012) Morphocons: a new sonification concept based on morphological earcons. J Audio Eng Soc (accepted 2012–01–19)

Parseihian G, Brilhault A, Dramas F (2010) NAVIG: an object localization system for the blind. Workshop on Multimodal Location Based Techniques for Extreme Navigation (Pervasive 2010), Helsinki, Finland

Ran L, Helal S, Moore (2004) S Drishti: an integrated indoor/outdoor blind navigation system and service. IEEE international conference on pervasive computing and communications (PerCom'04), pp 23–30

Roentgen UR, Gelderblom GJ, Soede M, de Witte LP (2008) Inventory of electronic mobility aids for persons with visual impairments: a literature review. J Vis Impair Blind 102(11):702–724

Shi J, Tomasi C (1994) Good features to track. Proc. CVPR'94. IEEE computer society conference on computer vision and pattern recognition, pp 593–600

Strothotte T, Petrie H, Johnson V Reichert L (1995) Mobic: user needs and preliminary design for a mobility aid for blind and elderly travelers. In: Porrero IP, de la Bellacasa RP (eds) The European context for assistive technology, pp 348–352

Thorpe S, Fize D, Marlot C (1996) Speed of processing in the human visual system. Nature 381(6582):520–522

Thrun S (2002) Robotic mapping: A survey. In: Lakemeyer G, Nebel B (eds) Exploring artificial intelligence in the new millennium, Elsevier Science, USA, pp 1–35

Tran T, Letowski T, Abouchacra K (2000) Evaluation of acoustic beacon characteristics for navigation tasks. Ergonomics 43(6):807–827

Vézien J-M, Ménélas B, Nelson J, Picinali L, Bourdot P, Ammi M, Katz BFG, Burkhardt JM, Pastur L, Lusseyran F (2009) Multisensory VR exploration for computer fluid dynamics in the CoRSAIRe project. Virtual Reality 13:257–271

Völkel T, Kühn R, Weber G (2008) Mobility impaired pedestrians are not cars: requirements for the annotation of geographical data. Computers Helping People with Special Needs, LNCS 2008 5105:1085–1092

Walker BN, Lindsay J (2005) Navigation performance in a virtual environment with bonephones. In: Proceedings of international conference on auditory display, Limerick, Ireland, pp 260–263

Walker BN, Lindsay J (2006) Navigation performance with a virtual auditory display: Navigation performance with a virtual auditory display: effects of beacon sound, capture radius, and practice. Hum Factors: J Hum Fact Ergonomics Soci Sum 48(2):265–278

Walker BN, Nance A, Lindsay J (2006) Spearcons: speech-based earcons improve navigation performance in auditory menus. In: Proceedings of international conference on auditory display (ICAD2006), pp 95–98

Zheng J, Winstanley A, Pan Z, Coveney S (2009) Spatial characteristics of walking areas for pedestrian navigation. Third International conference on multimedia and ubiquitous engineering, IEEE, China, pp 452–458