Mobile Networks and Applications manuscript No.
(will be inserted by the editor)

Spatialized audio environmental awareness for blind users with a smartphone

Jeffrey R. Blum · Mathieu Bouchard · Jeremy R. Cooperstock

Received: date / Accepted: date

Abstract Numerous projects have investigated assistive navigation technologies for the blind community, tackling challenges ranging from interface design to sensory substitution. However, none of these have successfully integrated what we consider to be the three factors necessary for a widely deployable system that delivers a rich experience of one's environment: implementation on a commodity device, use of a pre-existing worldwide point of interest (POI) database, and a means of rendering the environment that is superior to a naive playback of spoken text. Our "In Situ Audio Services" (ISAS) application responds to these needs, allowing users to explore an urban area without necessarily having a particular destination in mind. We describe the technical aspects of its implementation, user requirements, interface design, safety concerns, POI data source issues, and further requirements to make the system practical on a wider basis. Initial qualitative feedback from blind users is also discussed.

Keywords spatialized audio, blind navigation, GPS, smartphone, audio augmented reality

1 Introduction

Guide dogs and canes have long been the staple assistive devices used by the blind community when navigating city streets. More recently, GPS has broadened the possibilities for autonomous exploration. Efforts to date have largely been focused on guiding a user from one location to another, usually via turn-by-turn spoken directions, much as a car GPS system operates for drivers. A blind user can use such a system, although since the device and the database of geographic information are designed for automotive applications, they have significant limitations when used for pedestrian navigation by the visually impaired [25]. Devices designed specifically for the blind often run on custom hardware, and are for the most part single-purpose and relatively expensive, or else run on commodity hardware, but with significant limitations on functionality. Existing commercial audio-based tools typically rely exclusively on speech to indicate the distance, direction, and type of locations around a user. Unfortunately, this form of content delivery is intrusive or distracting, thus discouraging continuous use. Considerable research has been invested in using spatialized audio to navigate or render waypoints and points of interest (POI) information, but the resulting systems require the use of bulky, expensive, or custom hardware and are thus not well-suited for wide deployment. Many research systems also depend on proprietary POI databases that cover only a small area, and are not easy to generalize to multiple cities or countries. There has also been a significant body of work on augmented reality applications that create a virtual layer of additional content, rendered overtop of the real-world information. However, most such systems employ video overlays, which are inaccessible to the blind.

This work was made possible thanks to the financial support of the Québec Secrétariat du Conseil du trésor through the *Appui au passage à la société de l'information* program, as well as additional funding from a Google Faculty Research Award. This paper is an expansion and update of the conference paper *What's around me? Spatialized audio augmented reality for blind users with a smartphone*, presented at Mobiquitous 2011, Copenhagen, Denmark [2].

McGill University, Montréal, Québec, Canada
Tel.: +1-514-398-5992
E-mail: `{jeffbl,matju,jer}@cim.mcgill.ca`

The confluence of advanced smartphone technology and widely available geospatial databases offers the opportunity for a fundamentally different approach. The current generation of smartphones is sufficiently powerful to render multiple voices of spatialized audio, and also integrates GPS, compass, accelerometer and other sensors that allow for a complete audio augmented reality system that is useful and enriching to the blind community. However, despite tightly scoped efforts on the individual issues mentioned above, no attempts have been made to bring all of these pieces, including spatialization, together into a practical working application that can be used on an ongoing basis. Our objective is to create a solution usable by simply installing a piece of software on a widely available device, without additional hardware beyond the phone and a pair of headphones, and without depending on customized databases. Although this significantly limits the achievable accuracy, reliability and functionality of the resulting system, it potentially allows for more widespread use. Designing the system to be useful despite these constraints was our primary challenge.

In this context, we describe our smartphone application, In Situ Audio Services (ISAS). Rather than navigation assistance, ISAS exploits the resources noted above to provide the blind community with a greater sense of environmental awareness. Specifically, ISAS enables blind users to sense locations in their vicinity such as restaurants and shops. Our key criterion for success is not enabling users to navigate themselves through the door of a new restaurant without assistance, but rather walk down the street and serendipitously *notice* a restaurant of which they were previously unaware.

Although ISAS is specifically designed for the blind and vision-impaired community, the implications of this system extend beyond. Anyone whose eyes are occupied with other tasks, driving being an obvious example, could make use of many of the ideas and implementations described in this paper. We expect that in the future, much of the functionality we describe here will be adopted in various forms by the general population.

The contributions described by this paper center on our effort to combine three key items to form a practical audio augmented reality system for blind users: a readily available and unmodified smartphone platform, a commercial location database, and an audio spatialization implementation for orientation awareness. These are coupled with a novel user interface that addresses the needs of someone walking in an urban environment while simultaneously holding a cane or guide dog harness. Our implementation thus represents a "snapshot in time" that tests whether current platforms and POI services are indeed sufficiently capable to fulfil the long-standing vision of a small, mobile platform for such exploration, and if not, what specific issues remain before such a system can be practically deployed.

2 Previous Work

Numerous commercial systems exist for blind navigation, e.g., the HumanWare Trekker Breeze[1] and Mobile Geo,[2] the latter powered by Sendero GPS software,[3] which provide not only navigation assistance, but also POI information along the route. The free Loadstone GPS software for Nokia phones[4] allows users to import POI information and be informed when they approach the locations they have chosen to load into memory. Intersection Explorer[5] for Android smartphones lets users explore nearby streets and intersections by dragging their finger on the phone's touchscreen. Other recent smartphone applications include AriadneGPS,[6] which helps users navigate waypoints and keep track of their current location by address, and also allows map exploration by dragging a finger on the screen. OnTheBus[7] assists blind users with public transit by guiding them to the nearest bus stop, then reading out stops along the way so the user knows when to exit the bus. BlindSquare[8] focuses on blind accessibility to POI information from the widely used FourSquare[9] service that allows users to find locations of interest, read reviews, and "check in" to let others know where they are. Further systems are listed in a recent literature review [19]. However, none of these tools utilize spatialized audio, despite demonstrations that for wayfinding, the cognitive load of spatialized content is lower than when using language, e.g., spoken "left" or "right" instructions [12]. For recent systems, this is generally not due to hardware limitations, since some spatialized rendering capability is often included on the device for gaming applications, e.g., a limited OpenAL library on the iPhone. Although the iPhone implementation appears to be quite limited, a full head-related transfer function (HRTF) implementation for rendering 3D sound is built in to Nokia N95 phones, and has been shown to be effective [24].

Spatialized audio has, however, been employed in previous research systems. An early example is the Personal Guidance System (PGS), which used a GPS, com-

---

[1] `http://www.humanware.com`

[2] `http://www.codefactory.es/en/products.asp?id=336`

[3] `www.senderogroup.com/products/GPS/allgps.htm`

[4] `http://www.loadstone-gps.com`

[5] `http://tinyurl.com/IntersectionExplorer`

[6] `http://www.ariadnegps.eu`

[7] `http://www.onthebus-project.com`

[8] `http://blindsquare.com`

[9] `http://foursquare.com`

pass and spatialized speech to guide blind users by rendering nearby waypoints and POI information, either organized by proximity or presented in a clockwise fashion around the user [8, 7]. Experiments using this system were conducted on the University of California Santa Barbara campus, where the research team had access to highly detailed map information and a nearby differential GPS base station 20 km from their site [7]. Although we only became aware of the PGS system well into our design of ISAS, it is remarkable how many of the design decisions and much of the functionality were reproduced across the two systems. The benefit of time, of course, is that improvements and integration of technology allow for greater capabilities in a small package, while the current ubiquity of geospatial databases now provides access to a rich set of continuously updated POI information. Such resources were not available during the development of PGS, but its designers clearly paved the way to important follow-on research through their pioneering efforts. In this regard, we believe that our main contribution lies in the combination of the implementation on a small, portable, consumer device, and its mechanisms for linking into a worldwide database of points of interest. Moreover, our user interface has, from the outset, been advanced from a strongly user-centered design basis, relying on many iterations of design, testing, and refinement to best understand the needs and preferences of our target population. Similar systems followed PGS, albeit lacking spatialized audio, including Drishti [18, 9] and MoBIC [17, 16].

The SWAN project continues in the vein of PGS, enabling experimentation with rendering an audio scene while a user moves in the real world. To provide full functionality, SWAN requires add-on external devices such as a head-mounted inertial sensor or digital compass for orientation sensing. Using a portable Windows PC as its platform, SWAN can render spatialized sound using a full HRTF via OpenAL, while still fitting into a relatively small shoulder bag [29]. The user interface relies on a combination of speech recognition and audio menus, and is primarily targeted at waypoint finding and navigation, although it also supports user annotation of locations. POI support is mentioned, but it is unclear what sort of database is used and how many locations are covered. Similar multi-component hardware platforms have been used to render spatialized audio for environmental awareness [22]. The trend has been toward smaller and lighter systems as technology improves [14]. More recently, the NAVIG project [11], currently implemented on a laptop, not only includes spatialized audio rendering, but also binocular cameras mounted on a helmet, which are used to improve location accuracy and to recognize and help the user find objects in their environment. It has undergone preliminary testing with visually imparied users, and more formal user testing is anticipated for fall 2012. The NAVIG project also proposed an augmented geographic information system (GIS) optimized for blind users [10].

Other efforts modify the environment itself by distributing IR transmitters or radio receivers over an area of interest.[10] The Talking Points research project started with distributing RFID tags in the environment, then moved to distributed Bluetooth beacons communicating with a portable computer to render POI information [20]. However, the most recent version of their system, TP3, recently switched to WiFi, GPS, and compass sensors on a smartphone for finding the user's location and orientation [30]. Talking Points 3 has many similarities to our project, especially in fostering "serendiptious discovery" through rendering POI information via a smartphone. However, the TP3 system does not use spatialized audio, and is currently restricted to a manually created local POI database, testing 68 POIs in their study with blind participants. Further, their user testing was conducted indoors, using a "Wizard of Oz" method that allowed them to set the participant's position and orientation manually, and therefore did not rely on the actual device sensors.

More generally, researchers have attempted to sonify information usually represented via visual maps, tables and charts, or images. The most direct way to do this is simply to speak a description, but this approach quickly becomes tedious. Thus, efforts have focused on using techniques such as abstract musical tones (earcons [1]), sped up spoken text (spearcons [27]), and recorded sounds (auditory icons [6]), to represent a thing or idea. These techniques can reduce the time spent representing information to a blind user, as well as provide additional features such as an overview of large data sets or *gist* [32]. Of recent note, the Timbremap project uses an iPhone and stereo sonification to allow users to learn an indoor map by providing audio cues that convey different shapes [21]. Although focused on exploration rather than navigation, users are always expected to stop and focus entirely on the application, rather than passively experiencing ambient information about their surroundings while walking.

3 The ISAS Application

Noting the lack of an attempt to merge the best features of existing research platforms, such as spatialized

---

[10] Two examples are `http://talkingsigns.com` and `http://eo-guidage.com`.

audio and auditory icons, with the low cost and ubiquity of commodity smartphone devices, as well as with commercial POI databases, we developed the ISAS application. This involved four main challenges addressed in this section: rendering a spatialized audio scene; designing a practical user interface; compensating for, or at least degrading gracefully in the face of, unreliable GPS and orientation sensors; and relying on existing large-scale, but imperfect, location data sources.

Thus, our goal was to create a system that can serve as a prototype for an application that is fully functional when installed on a standard smartphone. We recognized immediately that a commodity smartphone would restrict the accuracy of the sensors and data, as well as the processing power available for rendering the audio scene. The question we wanted to answer was whether a useful system could still be implemented despite these constraints, or else identify the specific issues still blocking such a system from practical deployment.

3.1 Hardware

Due to its audio capabilities, position and location sensors, powerful CPU and blind accessibility features (VoiceOver), our initial implementation runs on the Apple iPhone 4. The only external hardware required by the system is a pair of headphones. Given the importance to blind users of unobstructed audio of their environment, e.g., traffic sounds, we recommend either open ear AirDrive headphones,[11] whose speakers are placed in front of each ear, or bone conduction headphones,[12] which rest on the bone in front of the ear. Admittedly, the sound quality of these technologies is inferior to over-the-ear solutions, and positioning the AirDrive headphones correctly is initially quite difficult, thus impacting the quality of spatialization. Despite these drawbacks, the trade-off for safety reasons is clearly worthwhile. We note with interest that others are not only successfully using bone conduction headphones for similar purposes [26], but are also discussing solutions for overcoming their limitations when used for spatialization [28].

Of course, some of the standard hardware on a smartphone, most notably the high-resolution display, is useless to a completely blind user. However, from a development perspective, the screen was invaluable while designing, testing and debugging ISAS, as we received rich real-time visual feedback on sensor performance and device status while walking outdoors. Although in theory, a custom device could be less expensive to produce by omitting irrelevant hardware, the market for blind products is much smaller than that for a general smartphone. For instance, the World Health Organization estimates that there are approximately 39 million blind people in the world,[13] whereas Apple sold over 37 million iPhones during their fiscal first quarter of 2012,[14] or almost enough to cover all the blind people in the entire world. Likely due to this enormous disparity in economies of scale, we do not know of a less expensive platform that is as capable as a commodity smartphone. In fact, custom hardware for blind users is sometimes priced much higher than an Apple iPhone, despite lacking a display. Specialized hardware such as the Kapsys Kapten,[15] designed to be smaller and somewhat less expensive than an unsubsidized smartphone, suffers from similar issues of sensor reliability and battery life constraints.[16] As a review of the Kapten from the American Federation for the Blind points out, “Users of the iPhone and other smartphones may not need a dedicated GPS device, as these products offer more convenient and less expensive GPS solutions” [4].

For blind users who have already invested in a smartphone for its built-in communication and organization capabilities, adding an application such as ISAS is likely to be much more cost effective than purchasing a secondary device. Perhaps even more important is the question of convenience. For instance, one of the blind participants in our studies who owns both a Trekker and an iPhone observed that he rarely uses the former but always carries his iPhone. Several participants commented on the similarities of ISAS and their Trekker devices, but noted that ISAS allows one to solicit additional details about points of interest. Although ISAS does not currently take advantage of it, a display could also be made useful for visually impaired, as opposed to completely blind, users.

3.2 Audio Scene Rendering and Spatialization

The iPhone’s SDK includes an OpenAL implementation, but it appears that its 3D sound spatialization falls back to simple stereo panning, lacking the robustness of a full HRTF implementation, which attempts to model how the human anatomy, in particular, the head and external ears (pinna), filter sound before it

---

[11] `http://www.airdrives.com`

[12] `http://www.audioboneheadphones.com`

[13] `http://www.who.int/mediacentre/factsheets/fs282/en/index.html`

[14] `http://www.apple.com/pr/library/2012/01/24Apple-Reports-First-Quarter-Results.html`

[15] `http://www.kapsys.com`

[16] `http://www.senderogroup.com/products/gps/kaptenreview.htm`

reaches the eardrums. One recent attempt to improve the iPhone's built-in OpenAL 3D sound support was inconclusive in its benefits [13]. We briefly attempted to use the open-source earplug~ HRTF implementation,[17] but found that even a single voice consumed prohibitive amounts of CPU. Thus, we decided to improve on the built-in spatialization, but not attempt a full HRTF implementation. To do this, we use the libpd[18] implementation of PureData (Pd) that runs on many platforms including smartphones. This allowed us to create a Pd patch that uses not only simple panning techniques, but also interaural time difference and filtering effects to spatialize the sound. A low-pass filter helps to distinguish locations in back of the user, which results in the sounds being more "muffled" and somewhat quieter as they move further behind the user's head. Volume falls off depending on the distance of the item from the user's current position. The application can render up to four simultaneous spatialized items; when an additional one is requested, the first is stopped. Note that we expect built-in 3D sound rendering to improve rapidly on mobile devices in the near future, primarily driven by increases in processing power and the demands of gaming applications. As noted previously, technical feasibility of HRTF implementations has already been demonstrated on the Nokia N95.

The user interface, described in the next section, renders up to three audio representations of a location depending on user preference and current mode:

- Spatialized *category name:* A spoken pre-recorded category name rendered by a text-to-speech (TTS) system, e.g., "restaurant," "shop," or "cafe."
- Spatialized *category audio icon:* A short sound, e.g., ice clinking in a glass representing a bar, or a drum beat for entertainment.
- Non-spatialized *details:* The full name of the location, usually with spoken confirmation of bearing and distance.

Audio icons have several advantages over spoken words. In particular, we expected that walking mode, described below, would benefit from shorter, clearer indications of locations surrounding the user. There is a trade off, however, against the greater specificity possible when using words. For example, a verbal category name for a restaurant can easily differentiate between fast food, a food court, and a normal restaurant. We conducted two informal walks, each with a blind participant, during which they qualitatively compared versions that used audio icons vs. spoken category names. Their feedback, coupled with multiple informal trials by the ISAS development team, led to the conclusion that designing icons to represent location categories is surprisingly difficult, as they must satisfy multiple criteria:

1. They must be easily distinguishable from real environmental sounds. For example, a siren sound for a fire station risks confusion with a real siren, thus connoting danger rather than safety.
2. A strong onset (attack) helps disambiguate one long sound from overlapping repetitions, and avoids masking from other nearly simultaneous icons.
3. Since the icons may be filtered, e.g., when they are behind the user, they must contain a sufficient distribution of frequencies so that the effect of filtering is perceptible, differentiated from other filtering effects, and leaves the sound in a recognizable state.

3.3 Application Design

As noted earlier, our system is not intended for navigation assistance but for exploration and discovery within one's environment. This functionality is provided by two primary modes. *Walking mode* is designed to be used while users are walking down the street, and not actively interacting with the device until they notice something of interest. This can be viewed as a background voice, designed to be turned on continuously as a user walks down the street. *Stop & listen mode* is designed for actively searching the immediate vicinity.

These modes are intended to work together, allowing the user to notice something in walking mode, then transition seamlessly to obtain more details and carry out further exploration as desired. For example, just as a sighted person walking down a familiar street may notice a new sign by virtue of its bright colors, the *walking mode* provides a blind user with the possible auditory equivalent, that is, an unexpected sound icon on a route the user has traversed many times in the past. If this icon is of interest, the user can obtain details about the new location directly in walking mode, or else switch to *stop & listen mode* to explore the area in more detail.

3.3.1 Walking Mode

This mode is engaged when the device is kept in a vertical position (i.e., as if in a front shirt pocket, or hanging from a pouch around the user's neck). We implemented two different sound triggering mechanisms for walking mode, a *radar* sweep, which plays sound nodes sequentially in a clockwise sweep around the user's head, and a *shockwave* mechanism, which instead plays the sound

---

[17] `http://puredata.info/community/projects/software/earplug`

[18] `http://gitorious.org/pdlib`

nodes in order of distance, from near to far. Both are similar to those described in the earlier PGS system [7]. However, we repeat the playback continuously as long as the user remains in walking mode, whereas PGS appears to allow only manual triggering of sweeps.

The *radar* mechanism allows users to place the direction of nodes sequentially in clockwise order simply by the order in which they play. In addition, to reinforce the mental image of a circular "sweep" around the user's head, a short spatialized tick is played every 7.5^o, thus indicating the current position of the radar sweep. A more intense sound is played at the user's left, front and right sides to assist further in registration. These cues also indicate that walking mode is still active and operating.

In an earlier experiment [5], we tested variants of these two mechanisms, with subjects sitting in a chair, and concluded that there were few meaningful performance differences between them. However, in the shockwave representation, perception of bearing information is dependent entirely on the quality of spatialization. Recognizing this as a potential downside, we opted to carry out our actual user testing with the radar mode, rendering objects as far as 150 m from the user. Following initial user feedback, the distance cutoff was reduced to 70 m, and can be further limited to a maximum number of POIs (choosing the closest) per sweep. Achieving what one might consider an ideal rendering of the scene is highly challenging. As the user typically moves and rotates, new items come into range. We must thus decide between rendering more than the maximum number of items the user has specified, or else dropping potentially relevant information. The latter, however, risks incompleteness, since some POIs within the cutoff limit might never be rendered. Early versions of ISAS completed one 360^o revolution of the radar sweep every 12–18 seconds, and in dense areas with many locations, the sweep was dynamically slowed to avoid an overwhelming number of simultaneous sounds. This resulted in a disturbing, arrhythmic ticking, and the sweep was considered too rapid. This feedback prompted us to increase the sweep time to a constant 36 seconds per revolution, without any dynamic changes based on local POI density. To avoid cluttering the audio scene with objects the user has already passed, we only play locations to the front and sides. Although we arrived at this design decision at an early stage of our research, we subsequently discovered that PGS [7] suggested the exact same approach. While each location is always spatialized relative to the device's (and we assume, the user's) current orientation, the radar sweep itself progresses independently. For example, if the user rotates by 90^o during a sweep, the spatialization of POIs will also shift immediately by that amount, but the scene will not skip or repeat nodes in that 90^o region.

However, despite the drawbacks noted above, the shockwave representation may be better suited to the needs of our users. Because this approach renders POIs in order of proximity, from near to far, the ordering is often likely to correspond to degree of importance to the user. This recently prompted us to begin user tests of the shockwave rendering strategy. Both sweep types are now available as options, although the default choice remains to be determined.

To hear details in walking mode, the user touches a finger on the screen, which pauses the current sweep and begins playing additional information for the last location heard. Sliding the finger to the left allows the user to hear details for locations further back (counterclockwise) in the sweep. Again, this is similar in functionality to PGS [7]. Lifting the finger restarts the radar sweep from where it was paused. In order to allow easy access to the touchscreen when in walking mode, we use a neck pouch that hangs around the user's neck and holds the phone upside down, with the screen facing outwards away from the user's body. This makes it convenient to manipulate the screen without fumbling with the phone. The neck strap is long enough that users can also raise the phone to their ears to take phone calls, without having to remove the device from the carrier. The touchscreen works through the plastic pouch, allowing it to be used in inclement weather as well. Concern for the risk of theft of the device has arisen, especially when the screen is facing outward. We are currently investigating screen dimming to reduce the blatant visibility of the device, but expect that a more opaque holder that does not interfere with use of the touchscreen will still be necessary, since screen dimming alone would still let others easily identify the phone.

The walking mode has evolved considerably over the course of the project, largely in response to user testing. Most importantly, the type of information for each location and the manner in which it is played has undergone significant modification. Initially, the radar sweep only played the spatialized audio icon or category name of the locations, allowing the audio of two or more icons at similar bearings to overlap each other. This was later modified so that the full name of each location was included in the output. The change resulted from an initial round of usability testing, conducted with six blind participants, using ISAS to accomplish specific tasks on a street in Montreal. Four of the participants were blind from birth, one for approximately 40 years and another for approximately 20 years. All used a white cane on a regular basis. In this experiment, playback of

Fig. 1: Neck carrier holding the iPhone for hands-free use in an upside down orientation, facing outward, to provide easy access to the touchscreen. Device is in a plastic pouch that protects it from the elements, yet does not interfere with touch events. User is wearing Airdrive headphones that do not obstruct the ear canal, so is able to hear ambient noise, important for safety in a busy urban environment.

the location name, albeit unspatialized, was included as an option during the radar sweep, in addition to always playing the spatialized audio icon or category name. Users generally found the inclusion of the name to be useful, preferring this to hearing only the audio icon or category name. Referring to the inclusion of the location name, Participant 6 indicated, "I find this one better, I find it more at ease. . . it gives you more information, it is less frustrating. . . do not have to concentrate as much. . . is less physical (you do not have to tap constantly). . . find this one gives you more information." Despite the fact that pausing the sweep to read the name of each POI significantly increases the time of each sweep, this is now the default behavior.

Since walking mode repeats the sweep continuously, it can become not only distracting, but potentially dangerous when the user needs to perform a task such as crossing the street. For such situations, users expressed a strong desire to be able to silence ISAS completely on demand. In the original implementation, this could be done by holding the device parallel to the ground (activating Stop & Listen mode, described below), which does not render POI information until initiated manually by the user. However, this required the user to hold the device in an inconvenient orientation at the very moment they needed to focus on a new task. To ameliorate this, we implemented two changes. First, tipping the phone horizontal to the ground (entering stop & listen mode) then back down now serves as a toggle to activate or silence walking mode. Second, in the silent state, tapping the screen triggers a single sweep, and then returns to silence. This control not only improves safety, but also allows the use of ISAS without fear of constant interruption, for example, while having a conversation with someone. At the same time, on-demand rendering of the walking mode information remains easily accessible by a single tap. This trade-off between automatically pushing data to the user based on the changing device context, e.g., location, orientation, and surrounding POIs, which can be overwhelming, vs. having the user "pull" information manually, is an ongoing tension when designing mobile applications [3].

3.3.2 Stop & Listen Mode

When the user tips the device so it is parallel to the ground, it enters *stop & listen mode*. In this mode, the user can more actively explore the area around them.

Our first version of Stop & Listen mode was operated by running a thumb up and down the screen. In this implementation, the bottom of the screen represents nearby locations, and the top represents locations at some maximal distance, currently set to 70 m[19], as with walking mode. As users drag a finger, each sound node is played as it is crossed. Thus, users can "look" ahead of their current location in a more controlled way than in radar mode, since they actively control the triggering of the sound nodes. The device can then be tipped to the side while the finger or thumb remains on the screen to hear the same unspatialized details that can be heard in walking mode. If the user maintains the device tipped to the side, the details for the next closest four locations will also be read. This is particularly helpful in very dense areas, when it can be difficult to isolate a single location while dragging a finger. If desired, users can drag a finger up the left edge of the screen to increase the normal "lookahead" range by five times. Note that for continuity, the sweep in walking mode begins from the last location played in stop & listen mode when the user switches from the latter to the former.

In user testing, however, this manipulation proved to be difficult physically, since it involved both sliding a finger on the screen, as well as tipping the device for details. Thus, we have since implemented a simpler version of Stop & Listen mode. This allows the user to make a rightward swipe gesture on the screen to hear the location next furthest from their current position. Similarly, a leftward swipe can be used to go back one

---

[19] Earlier versions extended to 150 m.

location. Expanding on this concept, it was then logical to add a "you are here" feature, which is standard in devices such as the Humanware Trekker. A simple tap instead of a swipe in this mode reads the user's current location as a street address, followed by a summary of the scene, with a count of nearby POIs in each category. Although we have retained the original version for advanced users with finer motor control, the simplified mechanism is the default.

3.3.3 Sensor reliability

Accurate location and orientation information is key to implementing a working spatialized audio system. For the compass, magnetic interference is a significant issue. In ad-hoc testing, for example, we observed a repeatable difference of 30^o in the compass reading when bringing the device near a parked sport utility vehicle. In addition to the magnetic compass, the iPhone also includes a gyro-based sensor which provides device rotation readings, although they are not relative to the cardinal directions. Since the compass provides an accuracy estimate with each reading, we originally implemented a sensor fusion between the gyro and the compass in order to improve the reliability of the heading information. When ISAS started, it needed one good compass reading before it could function, since the compass is the only hardware sensor capable of finding north, which we require for calculating the bearing to surrounding POIs. Once this reading was obtained, we calibrated the gyro with this value. From then on, we used the calibrated gyro reading for all heading values. Every time a valid compass reading was received, the gyro was recalibrated to north, both to correct for gyro drift as well as to benefit from potentially more accurate compass information. When the compass lost its heading entirely or the reported accuracy was worse than $\pm 30^o$, the readings were ignored. Thus, once we had a single good compass heading, we could rely on the gyro to carry the system though periods where the compass was known to be unreliable.

After completing this implementation, however, user feedback indicated that POIs were frequently being rendered far from their correct bearing or location. In an effort to determine why, and under what circumstances this occurred, we conducted a more formal evaluation of the device sensors. Preliminary results indicate that although the gyro values, as obtained through the iPhone Core Motion API, are sometimes very reliable, they often exhibit a drift up to almost $3^o/s$ over sustained periods. This prompted us to stop using sensor fusion and instead, to rely solely on the compass itself. Although when the gyro is well-behaved, the fusion algorithm provides superior results, in the presence of drift, the quality of fused output can be *much* worse than without such fusion.

The former case is illustrated in Figure 2. In leg three of the walk, beginning at approximately 140 seconds, the compass reports an error over 30°, causing the algorithm to switch to the calibrated yaw value (blue) instead of the compass (cyan). Yaw was calibrated to north based on the last compass reading at a reported accuracy of 30° or better. Despite yaw drift of $-0.32°/s$ in this leg, indicated by the slope of the raw yaw values (dark green), the calibrated yaw error (yellow), exhibits a mean absolute error of 6.59° with standard deviation 4.48°. Overall, this is better than the compass error, shown in red, with mean absolute error of 9.84° and standard deviation of 7.31°. Near the end of this leg, the compass reports improved accuracy, so the algorithm switches back to accepting compass values, recalibrating the gyro to north each time a new compass value is received.

The latter case is illustrated in Figure 3. At approximately 120 seconds, the compass reports an error estimate of 35°, causing the fusion algorithm to switch to the iPhone's yaw value, indicated in blue. Again, this was calibrated to north based on the previous compass reading, while the reported error was lower. In this instance, the compass (cyan) then moves closer to the ground truth value (black), and performs well throughout the rest of the walk, despite its poor reported accuracy. However, the raw yaw in the fourth and fifth legs of the walk drifts at approximately $1.34°/s$ and $2.14°/s$ respectively, resulting in very poor fusion results. This can be seen by comparing the yaw error (yellow) with compass error (red), both of which indicate absolute differences from ground truth.

Interestingly, the Android platform uses the magnetometer in its calculation of device orientation by default. Starting with iOS version 5, this capability has been added as an option in the iPhone API as well.

Although we cannot entirely overcome the sensor limitations, we are designing the system to be maximally robust to sensor errors, in the manner suggested in previous work attempting to use unreliable GPS on handheld devices [15]. In our current implementation:

- If the GPS reports an error that exceeds a predetermined threshhold, synthesized speech explicitly informs the user regularly.
- By default, ISAS provides the relative orientation to a given POI, in the form of "Restaurant Lola Rosa, front left." However, when the compass accuracy degraded, early versions of our software would generate a spoken warning, which we found to be intrusive. Fortunately, we noted that even without a

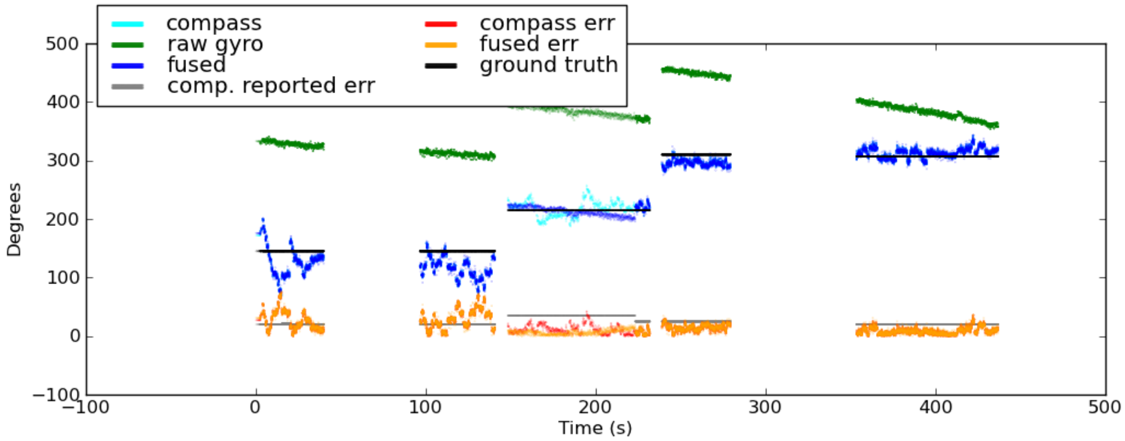

Fig. 2: Fusion algorithm working well: User is walking in five straight-line legs on a university campus and city streets.

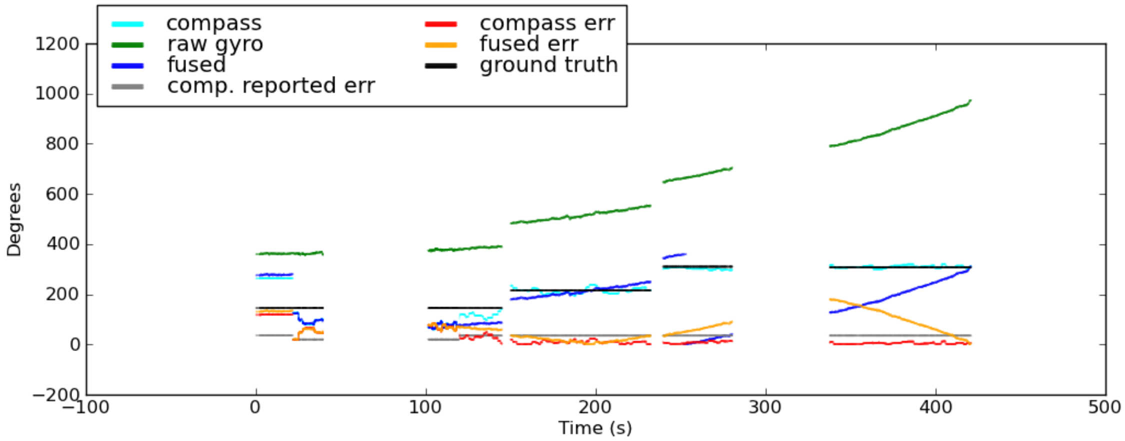

Fig. 3: Fusion algorithm working poorly.

compass, a reasonably accurate cardinal direction, e.g., “North” or “Southwest”, can be determined to a particular POI, as this depends only on the user’s current location. This allows us to provide some direction information, e.g., “Restaurant Lola Rosa, Southwest”, albeit only useful to users independently aware of their approximate orientation.

- Even when the GPS and compass readings are valid, we provide a low-level white noise in the background that varies in intensity based on their reported accuracy. This provides a continuous indicator analogous to the shaded GPS error circle visible in the iPhone’s map application.
- Locations within the current GPS error are rendered without spatialization (and thus currently sound like they are in front of the user), and the details report they are nearby, rather than providing a specific direction or distance.
- When describing the direction to a location, we use either clock directions (e.g., “9 o’clock”) or descriptions based on 45^o segments (e.g., “front” or “back left”), instead of a precise bearing in degrees, to avoid conveying more precision than warranted.

Thus, we depend on the sensors providing valid error estimates. Unfortunately, we see the same issue of poor error estimates from both the iPhone compass and GPS as documented for other mobile phone GPS implementations [31], with indications of reasonable accuracy when it is in fact off by 90^o or over a city block. When this occurs, we do not even know that the results are so inaccurate, and thus that nearby locations are rendered in the wrong directions. In the end, we must reinforce that ISAS does not replace a guide dog or cane, and there will be times when the information is simply inaccurate, without any indication. We expect this problem to fade as sensors improve, but it is a significant concern for deployment on currently available smartphone hardware, and reinforces the decision to provide a system for exploration rather than navigation.

3.3.4 Platform challenges

In addition to the hardware sensor issues described earlier, we have also encountered a number of practical challenges during our time implementing the ISAS software on the iPhone platform.

First, VoiceOver, which allows blind users to hear and manipulate all of the controls on the phone, is critical for practical adoption of the phone itself. However, since it filters all of the touch input to applications, it precludes the types of manipulation we have implemented in ISAS. Thus, on launch, we are in the position of having to tell the user to turn off VoiceOver in order to use ISAS, which is ironic given that ISAS is designed from the ground-up for blind use. Unfortunately, we have found no way to turn VoiceOver on and off automatically, nor even detect if it is currently enabled. Thus, we must resort to verbal instructions for turning it off on launch, and reminding the user to re-enable it when ISAS exits.

Second, the previous item is made additionally difficult since the user may switch to other applications for reminders, phone calls, or other system events. In iOS 5, a “Notification Center” was added, which scrolls down when touched to provide status of email, weather, or other items. This can be triggered accidentally when the user is manipulating ISAS, and with VoiceOver turned off, the user could become very confused. Since it is not possible to disable the Notification Center completely, nor do we want to disable other applications, we simply provide a spoken message whenever ISAS loses or gains focus, so users have some idea that the status of the application has changed, and they may need to toggle the VoiceOver mode.

Third, we are using the built-in iPhone text-to-speech engine, which makes ISAS ineligible for the Apple App Store, and thus limits its potential distribution. Switching to another TTS engine would resolve this issue.

3.4 Location data

In addition to relying on the sensors in the device, ISAS also depends on good location information. Other systems have used data sets generated specifically for their projects, or databases for small areas such as a university campus. This limits the area in which the tool is practically useful, and also imposes an ongoing maintenance burden. Increased interest in location-based service (LBS) products such as Foursquare has led providers including OpenStreetMap, Google Places and the Yellow Pages Group to offer application developers access to constantly updated databases of geographic content including businesses and other points of interest.

Initially, we were attracted to the OpenStreetMap service due to its openness and the ability to add our own data into the system. However, although the data was generally accurate, the quantity was insufficient. We now use the Google Places API to furnish data to the ISAS application, which provides an almost overwhelming amount of information on dense urban streets. Unfortunately, the way the data is generated causes some issues for ISAS.

A substantial portion of the latitude/longitude data appears to be generated by interpolating the location along the street using nothing more than the address to determine where along the street it lies. For example, the McCord Museum in Figure 4a is located on the opposite side of the sidewalk from the actual building, and past the building along the street. For car navigation, this is not a significant issue, as it is accurate to the city block, but for a blind pedestrian, this means that while walking on the sidewalk in a northeast direction, she would pass the actual museum on the right side, while hearing it spatialized off to the left. Then, immediately after passing the building, she would hear it directly to her left, despite it being to her back right. Further, the data is inconsistent, as in Figure 4b, which shows three restaurants all in the same building. One of them (the Subway restaurant) is placed exactly at the entrance door, which is the best possible case. Art Java is near enough to the entrance door that the error is not likely to be significant compared to GPS and compass errors on the device. However, the Broadway Cheesecake restaurant is placed on the street corner, far from the building, resulting in the same issue seen in Figure 4a. Fortunately, these issues subside somewhat in areas of the city where there are more discrete buildings with individual addresses, as in Figure 4c, although a significant portion of the data is still placed directly on the street.

To overcome these problems in part, we are now experimenting with “snapping” the user’s location to the middle of the closest street, as is frequently done in a car GPS unit, where the system can generally assume that the car is travelling on a roadway. Although we cannot necessarily make the same assumption when tracking a pedestrian, we expect that a similar algorithm can improve the quality of the rendered scene. For example, if users virtually “drift” into a building due to GPS error, the algorithm should move their position back to the correct side of the POI. This may be especially useful if a POI is placed directly on the

(a) (b) (c)

Fig. 4: Examples demonstrating Google Places data issues for pedestrian use.

edge of the street, as discussed above, since this results in rendering on the wrong side, even when the GPS is perfectly accurate.

Of course, one of the downsides of this approach is that the estimated distance to the POI may be less accurate. However, as was expected, user feedback indicated that rendering POIs on the incorrect side of the street was a far more serious problem, and greatly reduced confidence in the overall system. Other anticipated failure cases include the possibility of the snapping algorithm placing the user on the wrong street. To mitigate against serious errors of this type, it may be useful to introduce additional feedback, alerting users when the current street changes due to the snapping algorithm. This would allow for monitoring of the system accuracy, and for users to ignore the output when it appears they have been "teleported" to an incorrect location. Another approach that could be considered is to introduce heuristics that allow the system to snap the user to a different street only when the algorithm considers this to have occurred with some probability above a threshold. However, making such decisions with confidence becomes challenging in various scenarios, such as narrow city blocks and user movement through large open spaces such as parks.

Our "snapping" implementation uses street segment data from OpenStreetMap, which is downloaded and cached on the device based on its current location, similar to the other POI information. Once the user moves sufficiently to require another download, the new results are merged with previously downloaded content and data far outside of the user's current radius is pruned from the cache. The caching algorithm uses different update thresholds and query radii depending on the data; for example, we use a larger radius when querying the street segments necessary for the "snapping" algorithm than when querying for POI content. This ensures that there are enough segments in range to determine the nearest street accurately.

Whenever ISAS receives a location update, it uses the locally cached street segment data to calculate the latitude and longitude of the nearest point on the closest street segment. This point is determined by a brute-force trial of all possible linear projections to street segments, then used instead of the raw GPS coordinate to represent the user's location.

We have also integrated other frequently requested data sets into ISAS, specific to Montreal, such as bus stop and subway data. These are currently licensed only for research use, and thus cannot be deployed widely. We have not found a global source for this type of information, so this may need to be accomplished on a city-by-city basis for the near future.

Finally, location names are usually not tagged by language, so we cannot easily indicate the correct text-to-speech voice to use. In bilingual Montreal, it is difficult to understand the French voice speaking "Ye Olde Orchard Tavern" just as it is difficult to understand the English voice speaking "Lévesque".

3.4.1 Content server

Since we include data from multiple sources, we have also created a server system that queries or stores a local copy of the necessary information, then consolidates and sends it to the mobile client on request. This architecture, described below, is illustrated in Figure 5, and is currently broken down as follows:

1. Generic POI content, such as businesses and landmarks: queried by the ISAS server via the Google Places API on each device request, but not cached on the server. Since there is a limit to the number of POIs returned per Google query, multiple queries broken down into sets of categories are executed, then consolidated to ensure most of the POIs in the radius are returned. If the radius of the device query becomes too large, one central circle plus six surrounding circles are "probed." In this case, the

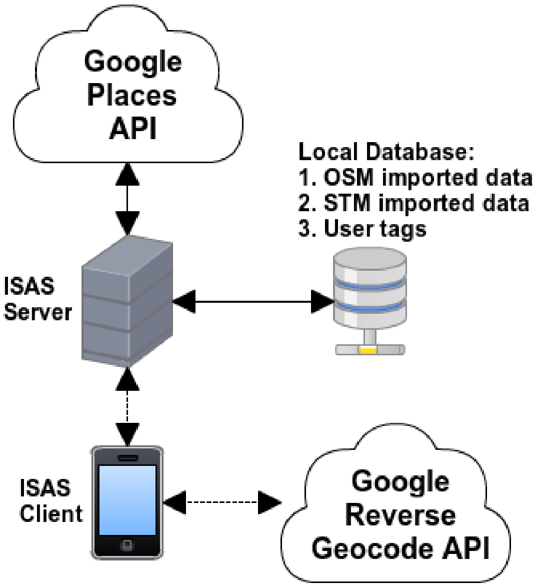

Fig. 5: ISAS content server architecture. Dashed lines indicate wireless connections. Note that for efficiency, most data is proxied through the ISAS content server rather than directly queried from the mobile client.

ISAS server currently generates a total of 49 Google queries.
2. Bus and metro information for Montreal: obtained under a research license, and stored on the server. This is not updated frequently, as it was provided on physical media directly from the Société de transport de Montréal (STM). This data is password-protected to ensure that it is only accessible by users in our studies.
3. Street intersections and segments: downloaded from OpenStreetMap and stored on the server. This can be updated through a series of scripts that organize the raw data into a format optimized for our queries.
4. Polygon POIs: For large buildings, parks, and other open areas, using only a single latitude/longitude location is misleading. For example, if the user is standing at the entrance to a park, but the corresponding POI is represented by a point in the center, or worse, at the opposite end, the system would indicate that the park is far away. To address this problem, we have imported polygon data for all of Quebec from OpenStreetMap. We use this to calculate the nearest point on the polygon to the user's current location, and to determine whether the user is inside or outside of the polygonal region, using a point-in-polygon algorithm [23].
5. User-generated POIs: Although not yet tested with users, we have prototyped the ability to upload recordings from a device to create new POIs. These are stored only on our ISAS server.
6. Address information: a separate query to the Google reverse-geocode API is sent each time the device receives a location update, in order to find and cache the nearest address, in case the user requests it in stop & listen mode.

After each query from the device, the timing of which is determined by user movement, data obtained from all of these sources is rendered in an XML format, compressed, and sent to the device. User-generated audio recordings are downloaded separately via a link in the associated XML POI. No state is kept on the server regarding individual devices; each phone is responsible for determining when it needs new data. This is non-optimal since the same POIs can be sent in multiple queries, due to overlap in the regions for each query. Currently, the duplicate nodes are simply discarded when received on the client. Although this is currently not a serious issue, in the future, bandwidth wastage could be reduced by a smarter query algorithm or retention of some client state on the server.

4 User Feedback

Quantitatively evaluating a system like ISAS, which seeks to improve the quality of life for blind users, is difficult. So far, we have carried out both informal usability evaluations with sighted team members and two blind participants walking in downtown Montreal, followed by two more formal studies, each with six blind participants. Several important findings came out of the feedback from these sessions.

First, there was an insistence that the system not "get in the way" by playing long segments of audio that cannot be interrupted. This makes sense, as audio is largely a linear communication method, unlike vision, with which we can more easily make sense of multiple inputs at the same time. Thus, ISAS spatialized categories and audio icons are kept short, and the user must explicitly request longer details. Even then, the user can immediately cancel the details with a simple gesture. This decision encourages the user to get details frequently, since the cost of cancelling the operation is minimal. In addition, keeping the communication terse is crucial to a good experience. Aside from carefully wording messages to be as short as possible, we also round distance values as they are further from the user, since "one hundred forty seven meters" is longer than a rounded "one hundred fifty meters," and in French, "quatre-vingt-dix-neuf" (99) is significantly longer than "cent" (100).

Not surprisingly, our experience confirmed that feedback obtained in a laboratory environment is often of limited value, or worse, occasionally misleading. Sitting in a quiet environment listening to test versions of the software, we would frequently overrate the amount of information a user could handle. These naive estimates were quickly dispelled after trying the system on a real (noisy) street. This pitfall not only applied to the sighted members of the ISAS team. For example, a blind user tried the system in a quiet indoor environment, and indicated it would be good if the details for a location would be spatialized along with the category indicator. However, at a later date, while using the system outdoors and walking down a busy street with a guide dog harness in one hand and ISAS in the other, the same user explicitly indicated it was good that the details were not spatialized since it was more important that the details cut through all of the distractions and surrounding noise.

Accompanying blind users while they use ISAS has illustrated several other interesting effects. For example, one user heard the actual sounds of people eating on a restaurant terrace. He commented that he would like to know the restaurant's name, so he entered *stop & listen* mode, pointed the device in the direction of the (real) sound, and attempted to find the restaurant. Unfortunately, it was not in the database at that time, but we note that the combination of real-world sounds with the augmented reality provided by ISAS can form a powerful combination. This same blind user, while in walking mode, paused to get details on a restaurant, then while gesturing in the correct direction, mentioned he did not know there was a cheesecake restaurant over there. These are the types of serendipitous discovery we hope ISAS will enable for more users.

After completing the first user test, we questioned the ecological validity of a short-term experiment outside of the user's usual context, compounded by the presence of an accompanying experimenter. To address these concerns, we completed a set of longer-term trials in which six participants were loaned a device to use in their daily routine for a week or more. This allowed them to provide more informative feedback regarding their experiences with the system within their daily routine. Feedback pointed to user excitement for ISAS, albeit with two important problems that we have since attempted to address, to be verified in the next round of user testing. First, users frequently felt overwhelmed by the quantity of information provided in walking mode, resulting in our modifications to limit such audio output. Moreover, as described earlier, we now allow the walking mode to be toggled on and off easily, and provide more manual control for the triggering of sweeps while walking. Second, participants strongly prefer to avoid having to hold the device continuously while walking. This prompted the fabrication of a harness (described earlier) that supports the smartphone in a stable position while allowing easy manual interaction with the device. Similarly, it also motivated us to ensure that the device need not be held in a particular position to keep it silent.

Throughout the development, we have emphasized experimentation with users in natural settings. Applying the feedback gained from these sessions has been a significant factor in the evolution of our interaction design, and one, we believe, that may distinguish the user experience with ISAS from that of other systems. For example, Ariadne GPS opts for a map-based representation in which the top-half of the smartphone screen includes locations in front, the lower-half represents locations behind, and the center of the screen corresponds to the user's current location. Ariadne GPS allows users to point the device in a desired direction or employ finger-based exploration of the map, an obvious navigation metaphor for sighted users. However, our experience suggests that the latter raises problems for the visually impaired, not the least of which is the difficulty of establishing the center point of the screen in the absence of a tangible cue or remembering the precise screen location corresponding to another point [5]. Moreover, our testing indicated that users typically referred to objects and points of interest ahead of them. In fact, our participants often found it confusing or annoying to hear information to related points of interest they had already passed. This led to our limiting the ISAS radar sweep to exclude content behind the user, as noted in Section 3.3.1. In general, our approach favours the use of generic gestures, e.g., flipping the phone up to enter *Stop & listen mode* and swiping left or right to proceed backward and forward through the details of nearby locations.

In a similar vein, our approach differs from BlindSquare, which, like ISAS, reports on locations of interest in the user's vicinity and employs a similar breakdown of content by categories as used in ISAS. The most significant differences are our exploitation of spatialized audio for rendering an ongoing display of the environment in a geographically organized manner (i.e., by direction or distance), as well as the option of rendering information regarding points of interest based on their position relative to a nearby intersection. In contrast, BlindSquare reads POIs filtered by their FourSquare popularity and organized by category (e.g., food, nightlife spot, shop & service), and always in an egocentric manner. Although BlindSquare does read popular POIs at intervals while walking, there is limited user control of this rendering beyond selecting categories and range. BlindSquare also does not utilize gestures with the device itself beyond shaking the phone to trigger reading the current location (address, direction, nearest intersection) information, or checking in on FourSquare. In contrast, ISAS uses the device's orientation sensors to change modes and general touchscreen gestures to move between POIs.

5 Future Work

Our ongoing work includes refinement of the accuracy and reliability of the system and porting to other platforms, e.g., Android, which is now operational. We are also completing the implementation of features such as user tagging, where users can record new POIs based on their current location, which are uploaded to our server and provided to future visitors to the same location. We expect this to be most useful for identifying hazards or other information most useful to blind users, as a supplement to the generic location information provided by existing online POI databases. We have also not yet explored filtering of content by category or other criteria to reduce the overwhelming nature of areas with many POIs nearby, crucial for further reducing the intrusiveness of the system in walking mode. Although not the focus of this work, we have also considered adding support for optional hardware such as a head-mounted compass, which may not only be more reliable than the built-in smartphone sensor, but also allow the user to orient their head independently of the smartphone to better localize sounds. Last, we have not optimized ISAS for power consumption. Currently, we run the sensors continuously at a high refresh rate, but could be much smarter about reducing the rate or accuracy requirements. Even so, we anticipate that ISAS will typically be run in bursts throughout the day, potentially allowing for charging in between uses. For users who might require longer periods of uninterrupted use, many smartphone battery extension sleeves are also available.

6 Conclusion

We created a smartphone application that uses spatialized audio to render nearby locations to blind users as they walk down the street. We accomplished this using only built-in sensors to obtain location and orientation information, allowing it to be installed on any iPhone 4. Initial feedback indicates the system is promising as a practical tool. However, limitations in the iPhone hardware sensors and currently available location databases mean the system not only fails in some cases, but cannot always know, and therefore indicate, significant errors in the presented information. We conclude that given current smartphone capabilities, a practical system is possible only for non-critical exploration of an environment, and not for high-accuracy navigation tasks. We have thus focused ISAS on exactly these use cases. Further user testing, especially over longer periods of time, will reveal whether blind users find the system useful on an ongoing basis, or whether the issues we have discussed are too great to overcome in practice. Toward this goal, work has been carried out to refine the ISAS user interface, based on user feedback pointing out several areas that need improvement. Each such change must then be verified through testing with blind users. Although only a snapshot in time, ISAS illustrates how close a practical system is, especially for a system targeted at exploration and ambient location information rather than navigation. Given the rapid rate of improvement in phone technology and geographic data sets, we have high hopes for such a system in the coming years. We expect that in the not-too-distant future, an application similar to ISAS will provide the blind community with an experience that matches the capabilities of research platforms such as SWAN, but implemented on a commodity smartphone, and deployed on a global scale.

Acknowledgments

The authors would like to thank Florian Grond, who implemented the original spatialization patches and consulted on audio issues throughout the project, Adriana Olmos, Dalia El-Shimy and Sabrina Panëels, who designed and carried out the user testing with blind participants, and Zack Settel and Mike Wozniewski. Stephane Doyon and Lucio D'Intino graciously provided feedback on the system while in development.

References

1. Meera Blattner, Denise Sumikawa, and Robert Greenberg. Earcons and Icons: Their Structure and Common Design Principles. *Human-Computer Interaction*, 4(1):11–44, March 1989.
2. J. Blum, M. Bouchard, and J. R. Cooperstock. What's around me? Spatialized audio augmented reality for blind users with a smartphone. In *Mobile and Ubiquitous Systems (Mobiquitous)*. Springer, December 2011.
3. Keith Cheverst, Keith Mitchell, and Nigel Davies. Exploring Context-aware Information Push. *Personal and Ubiquitous Computing*, 6(4):276–281, September 2002.
4. Jim Denham. Oh Kapten! My Kapten! Where am I?: A review of the Kapten PLUS Personal Navigation Device. *AccessWorld*, 12(7), July 2011.

5. Dalia El-Shimy, Florian Grond, Adriana Olmos, and Jeremy R. Cooperstock. Eyes-free environmental awareness for navigation. *Journal on Multimodal User Interfaces*, November 2011.
6. William Gaver. Auditory Icons: Using Sound in Computer Interfaces. *Human-Computer Interaction*, 2(2):167–177, June 1986.
7. R.G. Golledge, R.L. Klatzky, J.M. Loomis, J. Speigle, and J. Tietz. A geographical information system for a GPS based personal guidance system. *International Journal of Geographical Information Science*, 12(7):727–749, 1998.
8. R.G. Golledge, J.M. Loomis, R.L. Klatzky, Andreas Flury, and X.L. Yang. Designing a personal guidance system to aid navigation without sight: Progress on the GIS component. *International Journal of Geographical Information Science*, 5(4):373–395, 1991.
9. A.S. Helal, S.E. Moore, and B. Ramachandran. Drishti: an integrated navigation system for visually impaired and disabled. In *Proceedings Fifth International Symposium on Wearable Computers*, pages 149–156. IEEE Comput. Soc, 2001.
10. Slim Kammoun, Marc Macé, Bernard Oriola, and Christophe Jouffrais. Towards a Geographic Information System Facilitating Navigation of Visually Impaired Users. *Computers Helping People with Special Needs*, 7383:521–528, 2012.
11. Brian F. G. Katz, Slim Kammoun, Gaëtan Parseihian, Olivier Gutierrez, Adrien Brilhault, Malika Auvray, Philippe Truillet, Michel Denis, Simon Thorpe, and Christophe Jouffrais. NAVIG: augmented reality guidance system for the visually impaired. *Virtual Reality*, 16(3)(2006):1–17, June 2012.
12. Roberta L Klatzky, James R Marston, Nicholas A Giudice, Reginald G Golledge, and Jack M Loomis. Cognitive load of navigating without vision when guided by virtual sound versus spatial language. *Journal of experimental psychology. Applied*, 12(4):223–32, December 2006.
13. M. Linell. Comparison between two 3d-sound engines of the accuracy in determining the position of a source. Master's thesis, Luleå U. of Technology, 2011.
14. Nick Mariette. From Backpack to Handheld: The Recent Trajectory of Personal Location Aware Spatial Audio. *Media-Space Journal*, 1(1), 2008.
15. D. McGookin, Maya Gibbs, A.M. Nivala, and Stephen Brewster. Initial development of a PDA mobility aid for visually impaired people. *Human-Computer Interaction INTERACT 2007*, pages 665–668, 2007.
16. H. Petrie, V. Johnson, T. Strothotte, A. Raab, R. Michel, L. Reichert, and A. Schalt. MoBIC: An Aid to Increase the Independent Mobility of Blind Travellers. *British Journal of Visual Impairment*, 15(2):63–66, May 1997.
17. Helen Petrie, Valerie Johnson, Thomas Strothotte, Andreas Raab, Steffi Fritz, and Rainer Michel. MOBIC: Designing a Travel Aid for Blind and Elderly People. *Journal of Navigation*, 49(01):45–52, October 1996.
18. L. Ran, S. Helal, and S. Moore. Drishti: an integrated indoor/outdoor blind navigation system and service. *Second IEEE Annual Conference on Pervasive Computing and Communications, 2004. Proceedings of the*, pages 23–30, 2004.
19. Uta R. Roentgen, Gert Jan Gelderblom, Mathijs Soede, and Luc P. de Witte. Inventory of Electronic Mobility Aids for Persons with Visual Impairments: A Literature Review. *Visual Impairment and Blindness*, 102(11):702–724, 2008.
20. Jason Stewart, Sara Bauman, Michelle Escobar, Jakob Hilden, Kumud Bihani, and Mark W Newman. Accessible contextual information for urban orientation. In *Proceedings of the 10th international conference on Ubiquitous computing - UbiComp '08*, page 332, New York, New York, USA, 2008. ACM Press.
21. J. Su, A. Rosenzweig, A. Goel, E. de Lara, and K.N. Truong. Timbremap: enabling the visually-impaired to use maps on touch-enabled devices. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*, pages 17–26. ACM, 2010.
22. V. Sundareswaran, K. Wang, S. Chen, R. Behringer, J. McGee, C. Tam, and P. Zahorik. 3D audio augmented reality: implementation and experiments. *The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003. Proceedings.*, pages 296–297, 2006.
23. Evan E. Sutherland, Robert F. Sproull, and Robert A. Schumacker. A Characterization of Ten Hidden-Surface Algorithms. *ACM Computing Surveys*, 6(1):1–55, January 1974.
24. Yolanda Vazquez-Alvarez and Stephen Brewster. Investigating background & foreground interactions using spatial audio cues. *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems - CHI EA '09*, page 3823, 2009.
25. Thorsten Völkel, R. Kühn, and Gerhard Weber. Mobility impaired pedestrians are not cars: Requirements for the annotation of geographical data. *Computers Helping People with Special Needs*, pages 1085–1092, 2008.
26. B.N. Walker, J. Lindsay, and Others. Navigation performance in a virtual environment with bonephones. In *Proc. of the Int'l Conf. on Auditory Display (ICAD2005)*, pages 260–263. Citeseer, 2005.
27. B.N. Walker, A. Nance, and J. Lindsay. Spearcons: Speech-based earcons improve navigation performance in auditory menus. In *Proceedings of the International Conference on Auditory Display, London, UK*, pages 63–68. Citeseer, 2006.
28. B.N. Walker and R. Stanley. Thresholds of audibility for bone-conduction headsets. In *International Conference on Auditory Display, Limerick, Ireland*, pages 218–222, 2005.
29. Jeff Wilson, Bruce N. Walker, Jeffrey Lindsay, Craig Cambias, and Frank Dellaert. SWAN: System for Wearable Audio Navigation. *2007 11th IEEE International Symposium on Wearable Computers*, pages 1–8, October 2007.
30. Rayoung Yang, Sangmi Park, Sonali R. Mishra, Zhenan Hong, Clint Newsom, Hyeon Joo, Erik Hofer, and Mark W. Newman. Supporting spatial awareness and independent wayfinding for pedestrians with visual impairments. In *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility - ASSETS '11*, page 27, New York, New York, USA, 2011. ACM Press.
31. Paul a. Zandbergen and Sean J. Barbeau. Positional Accuracy of Assisted GPS Data from High-Sensitivity GPS-enabled Mobile Phones. *Journal of Navigation*, 64(03):381–399, June 2011.
32. Haixia Zhao, Catherine Plaisant, Ben Shneiderman, and R. Duraiswami. Sonification of geo-referenced data for auditory information seeking: Design principle and pilot study. *University of Maryland Tech. Report HCIL 2004*, 3:1–8, 2003.