What's around Me? Spatialized Audio Augmented Reality for Blind Users with a Smartphone

Jeffrey R. Blum, Mathieu Bouchard, and Jeremy R. Cooperstock

McGill University, Montréal, Québec, Canada
{jeffbl,matju,jer}@cim.mcgill.ca

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 preexisting 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 [19]. 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

A. Puiatti et al. (Eds.): MobiQuitous 2011, LNICST 104, pp. 49–62, 2012.

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.

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. 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 allows for potentially much more widespread use.

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 longstanding 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. Further systems are listed in a recent literature review [14]. 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 [7]. 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 [18].

Spatialized audio has, however, been employed in previous research systems. An early example is the Personal Guidance System (PGS), which used a GPS, compass 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 [5,4]. 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 [4]. Similar systems followed, albeit lacking spatialized audio, including Drishti [13,6] and MoBIC [11,12].

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 [23]. 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 [17]. The trend has been toward smaller and lighter systems as technology improves [9].

---

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

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

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

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

[5] `http://www.googlelabs.com/show_details?app_key=agtnbGFiczIwLXd3d3IVCxIMTGFic0FwcE1vZGVsGIHDtwIM`

Other efforts modify the environment itself by distributing IR transmitters, radio receivers[6] or RFID tags [15] over an area of interest.

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 [21]), and recorded sounds (auditory icons [3]), 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* [25]. 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 [16]. 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 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.

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 headphones,[7] whose speakers are placed in front of each ear, or bone conduction headphones,[8] 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

---

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

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

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

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 [20], but are also discussing solutions for overcoming their limitations when used for spatialization [22].

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 reaches the eardrums. One recent attempt to improve the iPhone's built-in OpenAL 3D sound support was inconclusive in its benefits [8]. We briefly attempted to use the open-source earplug~ HRTF implementation,[9] 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[10] 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

---

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

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

category name for a restaurant can easily differentiate between fast food, a food court, and a normal restaurant. Designing icons to represent location categories turned out to be 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.

Walking Mode. This mode is engaged when the device is kept in a vertical position (i.e., as if in a front shirt pocket). We implemented two different sound triggering mechanisms for walking mode, similar to those described in the earlier PGS system [4]. The *radar* mechanism plays sound nodes out to as far as 150 m from the user, sequentially in a clockwise sweep around the user's head. The *shockwave* mechanism instead plays the sounds in order from near to far. Unlike the PGS implementation, we repeat the playback continuously as long as the user remains in walking mode. In an earlier experiment [2], 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 a mobile context, with the user walking down the street, we found that a simple version of the radar mechanism seemed less intrusive and easier to comprehend.

One revolution of the radar sweep takes 12 seconds. However, in dense areas with many locations, the sweep is dynamically slowed to avoid an overwhelming number of simultaneous sounds. A short spatialized tick is played every 15^o to indicate where in the sweep the radar is currently playing, and a more intense sound is played at the cardinal points to assist in registration. These cues indicate that walking mode is still active and operating, they reinforce the spatialization by creating a circular "sweep" around the user's head, and they allow the user to anticipate where spatialized locations will next play. To avoid cluttering the audio scene with objects the user has already passed, we only play locations to the front and sides. 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. This ensures that user movement or swaying does not lead to repetitions or variations in playback rate.

To hear details in walking mode, the user touches a finger on the screen, which pauses the radar 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 (counter-clockwise) in the sweep. Lifting the finger restarts the radar sweep from where it was paused.

Stop and 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 by running their thumb up and down the screen. The bottom of the screen represents nearby locations, and the top represents locations 150 meters away. As the user drags their finger, they cross sound nodes, which are played as they are crossed. Thus, the user can "look" out 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 keeps the device tipped to the side, the details for the next closest four locations to the initial one 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. To "look" further, the left edge of the screen can be used in a similar fashion to "see" out to 750 m. Note that when the user switches from stop & listen mode to walking mode, walking mode starts its sweep with the last location played in stop & listen mode in order to provide some continuity between the two modes.

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 implemented a sensor fusion between the gyro and the compass in order to improve the reliability of the heading information. When ISAS starts, it needs one good compass reading before it can 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 is obtained, we calibrate the gyro with this value. From then on, we use the calibrated gyro reading for all heading values. Every time a valid compass reading is received, the gyro is recalibrated to north, both to correct for gyro drift as well as to benefit from potentially more accurate compass information. When the compass loses its heading entirely or the reported accuracy is worse than $\pm 30^o$, the readings are ignored. Thus, once we have a single good compass heading, we can rely on the gyro to carry the system though periods where the compass is known to be unreliable. Although we cannot overcome the sensor limitations entirely, we are designing the system to be as robust as possible to sensor errors, as suggested in previous work attempting to use unreliable GPS on handheld devices [10]. In our current implementation:

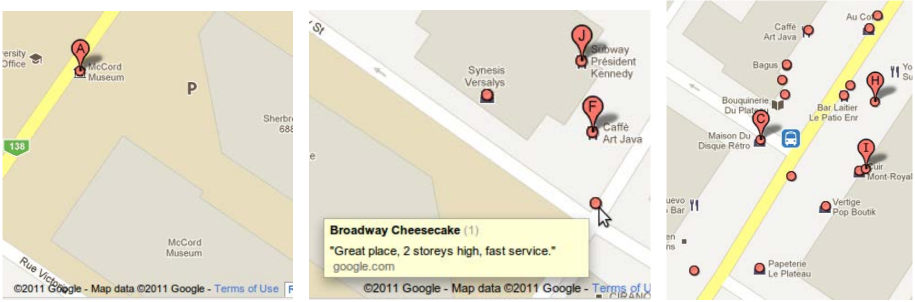
- If the GPS or the compass reports an error that exceeds a predetermined threshhold, synthesized speech explicitly informs the user regularly.
- Even when the GPS and compass readings are valid, we provide a low-level white noise in the background that varies in intensity based on their reported accuracy. 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., "1 o'clock" or "9 o'clock") or descriptions based on 45^o segments (e.g., "front" or "back left"), 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 [24], with indications of "good accuracy" when it is in fact off by 180^o or two city blocks. When this occurs, we do not even know that the results are 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 warning or other indication. We expect this problem to fade as sensors improve, but it is a significant concern for any deployment on currently available smartphone hardware, and reinforces the decision to provide a system for exploration rather than navigation.

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 the project, or databases for small areas such as a University campus, which limits the area in which the tool is practically useful, and also imposes an ongoing maintenance burden. Due to increasing interest in location based service (LBS) products such as Foursquare, providers including OpenStreetMap, Google Places, the Yellow Pages Group and SimpleGeo are offering 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



(a) (b) (c)

Fig. 1. Examples demonstrating Google Places data issues for pedestrian use

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 1a 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 1b, 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 1a. Fortunately, these issues subside somewhat in areas of the city where there are more discrete buildings with individual addresses, as in Figure 1c, although a significant portion of the data is still placed directly on the street.

In addition, location names are usually not tagged by language, so we cannot easily indicate the correct text-to-speech voice to use. In bilingual Montréal, it is difficult to understand the French voice speaking “Ye Olde Orchard Tavern.”

4 User Feedback

Quantitatively evaluating a system like ISAS, which seeks to improve the quality of life for blind users, is difficult. Although we are in the process of designing more formal tests, we have carried out informal usability evaluations with sighted team members and two blind participants walking in downtown Montréal. 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.

5 Future Work

We are currently conducting more formal user tests and revising the ISAS application based on the feedback being received. An initial round of usability testing involved six blind participants, using ISAS to accomplish specific tasks on a street in Montreal, with the sessions observed by a sighted experimenter. For this experiment, we also prototyped the ability to play the name of the location in walking (radar) mode, albeit unspatialized, in addition to the spatialized audio icon or category name. Users generally found this to be more useful than hearing only the audio icon or category name. For example, participant 6 indicated, "I find this one [with the location name read after the spatialized category] better, I find it more at ease... it gives you more information, it is less frustrating... you do not have to concentrate as much... and is less physical (you do not have to tap constantly)... I find this one gives you more information." However, the additional information spoken by the system results makes this method significantly more intrusive.

In addition, 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 are now in the midst of longer-term trials in which the participants are loaned a device to use in their daily routine for a week or more, allowing them to provide more informative feedback regarding their experiences with the system within their daily environment. Feedback received to date points to user excitement for ISAS, albeit with two important problems that we are currently addressing. First, users frequently felt overwhelmed by the quantity of information when using walking mode. Second, participants strongly prefer to avoid having to hold the device continuously while walking. This motivates our investigation of a suitable harness that supports the smartphone in a stable position while allowing easy manual interaction with the device.

Our ongoing work also includes refinement of the accuracy and reliability of the system and porting to other platforms such as Android. We are also adding 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. This is crucial, since initial usability testing and longer-term trials indicate that our current simple model of rendering points is too intrusive; making the walking mode "smarter" about what information it plays is our current priority. 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, there are several practical issues that need to be resolved. For example, we are currently 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. We are also integrating additional (and commonly requested) data sets into ISAS which are specific to Montreal, such as bus stop and subway data, which is currently only licensed for research use, and thus cannot be widely deployed. 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.

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. More work is also required to refine the ISAS user interface, as existing user feedback has pointed out several areas that need improvement. 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 truly global scale.

Acknowledgments. 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. The authors would like to give special thanks to Florian Grond, who implemented the original spatialization patches and consulted on numerous audio issues throughout the project, as well as Adriana Olmos, Dalia El-Shimy, Sabrina Panëels, Zack Settel, and Mike Wozniewski. Stephane Doyon and Lucio D'Intino graciously provided feedback on the system as it was in development.

References

1. Blattner, M., Sumikawa, D., Greenberg, R.: Earcons and icons: Their structure and common design principles. Human-Computer Interaction 4(1), 11–44 (1989)
2. El-Shimy, D., Grond, F., Olmos, A., Cooperstock, J.R.: Eyes-free environmental awareness for navigation. Springer Journal on Multimodal User Interfaces, Special Issue on Interactive Sonification (2011) (in press)

3. Gaver, W.: Auditory Icons: Using Sound in Computer Interfaces. Human-Computer Interaction 2(2), 167–177 (1986)
4. Golledge, R.G., Klatzky, R.L., Loomis, J.M., Speigle, J., Tietz, J.: A geographical information system for a GPS based personal guidance system. International Journal of Geographical Information Science 12(7), 727–749 (1998)
5. Golledge, R.G., Loomis, J.M., Klatzky, R.L., Flury, A., Yang, X.L.: Designing a personal guidance system to aid navigation without sight: Progress on the GIS component. Int. J. of Geographical Information Science 5(4), 373–395 (1991)
6. Helal, A.S., Moore, S.E., Ramachandran, B.: Drishti: An integrated navigation system for visually impaired and disabled. In: ISWC 2001 Proceedings of the 5th IEEE International Symposium on Wearable Computers (4514203), p. 149. IEEE Computer Society (2001)
7. Klatzky, R., Marston, J., Giudice, N., Golledge, R., Loomis, J.: Cognitive load of navigating without vision when guided by virtual sound versus spatial language. Journal of Experimental Psychology. Applied 12(4), 223–232 (2006)
8. Linell, M.: Comparison between two 3d-sound engines of the accuracy in determining the position of a source. Master's thesis, Luleå U. of Technology (2011)
9. Mariette, N.: From Backpack to Handheld: The Recent Trajectory of Personal Location Aware Spatial Audio. Media-Space Journal 1(1) (2008)
10. McGookin, D., Gibbs, M., Nivala, A.-M., Brewster, S.: Initial Development of a PDA Mobility Aid for Visually Impaired People. In: Baranauskas, C., Abascal, J., Barbosa, S.D.J. (eds.) INTERACT 2007. LNCS, vol. 4663, pp. 665–668. Springer, Heidelberg (2007)
11. Petrie, H., Johnson, V., Strothotte, T., Raab, A., Fritz, S., Michel, R.: MOBIC: Designing a Travel Aid for Blind and Elderly People. Journal of Navigation 49(01), 45–52 (1996)
12. Petrie, H., Johnson, V., Strothotte, T., Raab, A., Michel, R., Reichert, L., Schalt, A.: MoBIC: An Aid to Increase the Independent Mobility of Blind Travellers. British Journal of Visual Impairment 15(2), 63–66 (1997)
13. Ran, L., Helal, S., Moore, S.: Drishti: an integrated indoor/outdoor blind navigation system and service. In: Proceedings of the Second IEEE Annual Conference on Pervasive Computing and Communications, pp. 23–30 (2004)
14. Roentgen, U.R., Gelderblom, G.J., Soede, M., de Witte, L.P.: Inventory of electronic mobility aids for persons with visual impairments: A literature review. Journal of Visual Impairment & Blindness 102(11), 23 (2008)
15. Stewart, J., Bauman, S., Escobar, M., Hilden, J., Bihani, K., Newman, M.W.: Accessible contextual information for urban orientation. In: Proceedings of the 10th International Conference on Ubiquitous Computing, pp. 332–335. ACM (2008)
16. Su, J., Rosenzweig, A., Goel, A., de Lara, E., Truong, K.N.: 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, pp. 17–26. ACM (2010)
17. Sundareswaran, V., Wang, K., Chen, S., Behringer, R., McGee, J., Tam, C., Zahorik, P.: 3D audio augmented reality: implementation and experiments. In: Proceedings of the Second IEEE and ACM International Symposium on Mixed and Augmented Reality, pp. 296–297 (2003)
18. Vazquez-Alvarez, Y., Brewster, S.: Investigating background & foreground interactions using spatial audio cues. In: Extended Abstracts CHI 2009, p. 3823 (2009)

19. Völkel, T., Kühn, R., Weber, G.: Mobility Impaired Pedestrians Are Not Cars: Requirements for the Annotation of Geographical Data. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2008. LNCS, vol. 5105, pp. 1085–1092. Springer, Heidelberg (2008)
20. Walker, B.N., Lindsay, J.: Navigation performance in a virtual environment with bonephones. In: Proceedings of the International Conference on Auditory Display (ICAD 2005), vol. 3(3), pp. 1–26 (2005)
21. Walker, B.N., Nance, A., Lindsay, J.: Spearcons: Speech-based earcons improve navigation performance in auditory menus. In: Proceedings of the International Conference on Auditory Display, London, UK, pp. 63–68 (2006)
22. Walker, B.N., Stanley, R.: Thresholds of audibility for bone-conduction headsets. In: International Conference on Auditory Display, Limerick, Ireland, pp. 218–222 (2005)
23. Wilson, J., Walker, B., Lindsay, J., Cambias, C., Dellaert, F.: SWAN: System for Wearable Audio Navigation. In: 2007 11th IEEE International Symposium on Wearable Computers, pp. 1–8 (October 2007)
24. Zandbergen, P., Barbeau, S.: Positional Accuracy of Assisted GPS Data from High-Sensitivity GPS-enabled Mobile Phones. J. of Navigation 64(03), 381 (2011)
25. Zhao, H., Plaisant, C., Shneiderman, B., Duraiswami, R.: Sonification of geo-referenced data for auditory information seeking: Design principle and pilot study. In: Proceedings of ICAD 2004-Tenth Meeting of the International Conference on Auditory Display (2004)