Rendering the world to blind people via spatialized audio

Jeffrey R. Blum, Dalia El-Shimy, Mathieu Bouchard, Jeremy R. Cooperstock
McGill University
Shared Reality Laboratory
Montreal, Canada
[jeffbl,dalia,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 pre-existing worldwide point of interest (POI) database, and a means of rendering the environment that is superior to 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. This research note provides a brief background of the technical and user aspects of the current system, then describes results from a recent qualitative user study, as well as a discussion of likely future paths for the project.

Keywords
spatialized audio, blind navigation, 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. 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.

The confluence of advanced smartphone technology and widely available geospatial databases offers the opportunity for a fundamentally different approach. 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.

The background portions of this research note are based on an earlier paper describing ISAS [1]. Here we focus on the challenges we are currently attempting to overcome, based on a more recent user study than previously reported.

2. RELATED WORK
Numerous commercial systems exist for blind navigation, e.g., the HumanWare Trekker Breeze which provides not only navigation assistance, but also POI information along the route. Intersection Explorer[1] 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,[2] 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[3] 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[4] focuses on blind accessibility to POI information from the widely used FourSquare[5] 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 [10]. 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 [6].

Spatialized audio has, however, been employed in previous

[1]`http://tinyurl.com/IntersectionExplorer`
[2]`http://www.ariadnegps.eu`
[3]`http://www.onthebus-project.com`
[4]`http://blindsquare.com`
[5]`http://foursquare.com`

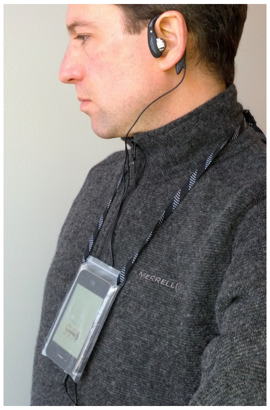

Figure 1: Neck carrier; open air headphones.

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 [3]. 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. Similar systems followed PGS, albeit lacking spatialized audio, including Drishti [9] and MoBIC [8]. The SWAN project continues in the vein of PGS, enabling experimentation with rendering an audio scene while a user moves in the real world [11]. More recently, the NAVIG project [5], 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. The NAVIG project also proposed an augmented geographic information system (GIS) optimized for blind users [4].

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. The current implementation runs on an iPhone or Android device. The phone is typically hung in a neck carrier around the neck, and is used with open air headphones that, for safety reasons, do not block the ear canal (Figure 1).

ISAS 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:* Full name of the location, usually with spoken confirmation of bearing and distance.

As noted earlier, our system is not intended for navigation assistance or obstacle avoidance, but for exploration and discovery within one's environment. This 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. [6]

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 nodes in order of distance, from near to far. Both are similar to those described in the earlier PGS system [3]. However, we repeat the playback continuously as long as the user is 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.

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 [3] suggested the exact same approach.

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 (counter-clockwise) in the sweep. Again, this is similar in functionality to PGS [3]. Lifting the finger restarts the radar sweep from where it was paused. For easy access to the touchscreen when in walking mode, we use a neck pouch that hangs around the user's neck.

Users generally found the inclusion of the spoken POI name to be useful, preferring this to hearing only the audio icon or category name, despite the fact that this makes each sweep take considerably more time.

Since walking mode repeats the sweep continuously, it can

[6]A demonstration video, including recorded spatialized audio, can be found at `http://isas.cim.mcgill.ca`

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. 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. 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.

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. This mode 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 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. We have also added an indication of where north lies, as well as the current reported accuracy of the compass and GPS.

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. For the GPS, tall buildings, clouds, and even position on the body can have significant effects. The reliability of these sensors was such an important consideration, and we had such trouble with both sensors in some cases, that we undertook a detailed quantitative study to assess when and how these sensors fail [2].

3.4 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, including: 1. POI content, such as businesses, via the Google Places API; 2. Bus and metro information for Montreal; 3. Street intersections and segments, from Open Street Map (OSM); 4. Polygon POIs, outlining such areas as parks that cannot be described by a single point, from OSM; 5. User-generated POIs: We have prototyped the ability to capture and upload recordings to create new POIs.

4. USER FEEDBACK

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, as well as several additional informal studies. Although many participants were excited about the idea of knowing what was around them from a system like ISAS, several important findings came out of the feedback from these sessions, which drove significant modifications to the system, as previously reported [7]. New to this research note, a qualitative test was run using the latest version of ISAS with three blind French-speaking participants in November 2012.

4.1 Methodology

Each of three blind participants, two male and one female, were played a spoken set of instructions, then walked individually with an experimenter on the same Montreal street with many shops and restaurants. None of the participants were more than passingly familiar with the area. All three had used earlier versions of ISAS and were familiar with iPhones. For the qualitative test, the participants were exposed to new features of ISAS, which included the ability to filter locations based on their categories, as well as three new options for rendering the position of POIs:

1. *Egocentric:* Direction and distance from user's position. "Front right, 30 meters"
2. *Intersection:* Position relative to a nearby intersection. "20 meters this side of intersection 2nd Avenue and Main Street"
3. *Dual:* Both of the previous methods combined. "Front right, 30 meters, 20 meters this side of intersection 2nd Avenue and Main Street"

Each participant was exposed to all three rendering options during specific portions of the walk. Category filtering was accomplished by the participants telling the experimenter that they wanted to filter by a certain category, and the experimenter manually setting ISAS to perform the desired filtering. Throughout the experiment, each participant was asked to "think out loud" and provide feedback about any aspect of the system. At various points during the walk, the experimenter solicited feedback on the features being tested, and made sure that all participants tried category filtering as well as the three different POI rendering strategies.

4.2 Results

Generally, filtering was perceived as useful for reducing the amount of information, as well as increasing its relevance. However, only one of the participants was overwhelmed by the amount of information being rendered without any filtering. The remaining two participants considered category filtering useful when they wanted to find only a specific type of location, such as restaurants. For POI position rendering, there was significant interest in the "dual" method. Further experimentation is required to validate this rendering.

However, for both the category filtering and the POI rendering method, there was a desire to be able to quickly switch between the options based on the current situation. For example, one participant indicated that having the "dual" rendering would be most useful in an unfamiliar area. One participant also reiterated feedback we had heard previously, wishing for turn-by-turn directions to a selected POI.

Overall, reaction from all three participants was positive. When participant #3 was asked if he would like to have ISAS on his own iPhone, he stated (translated), "I would love that! It would help me and I'd be able to test it, too." Further, he thought ISAS could potentially replace his Trekker Breeze.

5. FUTURE WORK

Our current priority is to move the use of ISAS beyond the limited number of users who have already tried it by releasing it in both the iPhone App Store as well as the Google Play Store. However, to support such a release, we must incorporate data for regions beyond Quebec on our server, and we need to integrate additional POI data sources; adding a data source like Foursquare would allow us to improve the relevance of the information via user ratings. Without this, we fear that even when users can easily toggling between categories (e.g., food, transportation, etc.), we will still be rendering too many low-value POIs, overwhelming the user.

We have also been working toward a collaboration with another University team to implement a better set of headphones, completely transparent to ambient sound while providing a higher quality audio signal from ISAS, potentially achieving much better spatialization and clarity.

Given sensor issues, we have considered support for an optional higher quality external GPS, or a head-mounted compass, which would also allow the user to orient their head independently of the smartphone to better localize sounds.

Based on ongoing feedback throughout the project, we are examining ways to hand off the location of a POI to an existing turn-by-turn navigation system to help guide users to the actual location once they are aware of it via ISAS.

Such changes, coupled with wider deployment, will allow us to take ISAS closer to a practical tool for blind users. However, we have also discussed releasing ISAS as an open-source platform for others to experiment with rendering geographic spatialized audio content not only to blind people, but anyone who might desire eyes-free awareness of their surroundings. The lack of such a common platform for experimentation, especially one that provides easy access to geographic data and spatialized audio rendering, makes it difficult for others to implement and test their ideas.

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. Upcoming work is focused on deploying ISAS to a wider user base for testing, which requires significant new feature work on both the client and server.

7. ACKNOWLEDGEMENTS

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

8. REFERENCES

[1] J. R. Blum, M. Bouchard, and J. R. Cooperstock. Spatialized audio environmental awareness for blind users with a smartphone. *Mobile Networks and Applications*, pages 1–15, December 2012.

[2] J. R. Blum, D. Greencorn, and J. R. Cooperstock. Smartphone sensor reliability for augmented reality applications. In *Mobile and Ubiquitous Systems (Mobiquitous)*, December 2012.

[3] R. Golledge, R. Klatzky, J. Loomis, J. Speigle, and J. Tietz. A geographical information system for a GPS based personal guidance system. *Intl. J. of Geographical Inf. Science*, 12(7):727–749, 1998.

[4] S. Kammoun, M. Macé, B. Oriola, and C. Jouffrais. Towards a Geographic Information System Facilitating Navigation of Visually Impaired Users. *Computers Helping People with Special Needs*, 7383:521–528, 2012.

[5] B. F. G. Katz, S. Kammoun, G. Parseihian, O. Gutierrez, A. Brilhault, M. Auvray, P. Truillet, M. Denis, S. Thorpe, and C. Jouffrais. NAVIG: augmented reality guidance system for the visually impaired. *Virtual Reality*, 16(3)(2006):1–17, June 2012.

[6] R. L. Klatzky, J. R. Marston, N. A. Giudice, R. G. Golledge, and J. 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, Dec. 2006.

[7] S. Panëels, A. Olmos, J. R. Blum, and J. R. Cooperstock. Listen to it yourself! evaluating usability of "what's around me?" for the blind (accepted). In *ACM Conference on Human Factors in Computing Systems (CHI)*, Paris, France, April 2013.

[8] 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.

[9] L. Ran, S. Helal, and S. Moore. Drishti: an integrated indoor/outdoor blind navigation system and service. *Proc. Second IEEE Annual Conference on Pervasive Computing and Communications*, pages 23–30, 2004.

[10] U. R. Roentgen, G. J. Gelderblom, M. Soede, and L. 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.

[11] J. Wilson, B. N. Walker, J. Lindsay, C. Cambias, and F. Dellaert. SWAN: System for Wearable Audio Navigation. *2007 11th IEEE International Symposium on Wearable Computers*, pages 1–8, Oct. 2007.